2018

Continuous Identity Verification in Cloud Computing Services

Al-Bayati, Burhan Mollan Salih

UNIVERSITY OF

PLYMOUTH

Continuous Identity Verification in Cloud Computing Services

by

Burhan Mollan Salih Al-Bayati

A thesis submitted to the University of Plymouth

in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Computing, Electronics and Mathematics

2018

COPYRIGHT STATMENT

Acknowledgements

I would like to begin by thanking ALLAH, without whose guidance this research would not have been possible to be accomplished. I would like to express my sincere gratitude to my Director of Studies Professor Nathan Clarke for the continuous support of my Ph.D. study and related research, for his patience, motivation, and immense knowledge. I am grateful to have a supervisor who cared so much about my work, and who responded to my questions and queries so promptly.

Thanks must also go to my other supervisors: Doctor Paul Dowland, Fudong Li who have spent a lot time proof reading papers and my thesis, in addition to providing helpful experience and guidance throughout my studies.

My sincere appreciation goes to the soul of my father (ALLAH have mercy upon him). I dedicated this thesis to him and my mother. I am very much indebted to my mother, brothers and sisters who supported during my PhD research journey.

Many thanks got to my wife Zainab who supported me in every day during my PhD research journey. She was very patient, understanding, and inspiring to me throughout this endeavour, spending days, nights, and sometimes even holidays without me. I should not forget to thank my heart, my son Abdullah and daughter Amna; May Allah bless them.

Special thanks also go to my best friend – Abdulwahid for his support and for the motivating ideas and thoughts, he provided during my PhD journey. I would also like to thank my colleagues who participated in my research namely: Abdulrahman Alruban and Hussain Alshamrani and all of my colleagues in the CSCAN department.

Finally, I would like to acknowledge with thanks and appreciation the government of Iraq, ministry of higher education and scientific research and university of Diyala, for granting me a scholarship and sponsoring my PhD studies. In addition, I would like to thank the Iraqi cultural attaché in London for supporting me during the period of study.

Author's Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Doctoral College Quality Sub-Committee.

Work submitted for this research degree at the University of Plymouth has not formed part of any other degree either at the University of Plymouth or at another establishment.
This study was financed with the aid of a scholarship from the government of Iraq.

Relevant seminars and conferences were attended at which work was often presented and several papers prepared for publication.

1. Al-Bayati, B., Clarke, N. and Dowland, P., 2016. Adaptive Behavioral Profiling for Identity Verification in Cloud Computing: A Model and Preliminary Analysis. *GSTF Journal on Computing (JoC)*, *ISSN: 2251-3043, Vol. 5, Iss.1, pp21-28, 2016*.
2. Al-Bayati, B., Clark, N., Haskell-Dowland, P. and Li, F., 2018, June. Continuous identity verification in cloud storage services using behavioural profiling. In 17th European Conference on Cyber Warfare and Security, ISBN: 978-1-911218-85-2, pp1-10. Academic Conferences and Publishing International Limited.
3. Al-Bayati, B., Clark, N., Haskell-Dowland, P. and Li, F., 2018. Misuse Detection in a Simulated IaaS Environment (in progress). It is accepted by Emerging Technologies for Authorization and Authentication (ETAA) 2018, and it will be published by Springer Verlag.

Word count of main body of thesis: 41,073 words

Signed.....

Date.....

Abstract

Continuous Identity Verification in Cloud Computing Services

Burhan Al-Bayati

Cloud computing has become a hugely popular new paradigm for hosting and delivering services over the internet for individuals and organisations with low cost. However, security is a sensitive issue in cloud computing, as it its services remain accessible to anyone after initial authenticated login and for significant periods. This has led to an increase in the number of attacks on sensitive customer information.

This research identified biometric approaches as a possible solution for security to be maintained beyond the point of entry. Specifically, behaviour profiling has been proposed and applied across various other applications in the area of Transparent Authentication Systems (TAS's) and Intrusion Detection Systems (IDS's) to detect account misuse. However, little research has sought to implement this technique within cloud computing services to detect misuse.

This research proposes a novel continuous identity verification system as a supporting factor to protect cloud users by operating transparently to detect abnormal access. The research examines the feasibility of applying a behavioural profiling technique on cloud users with respect to Software as a Service (SaaS) and Infrastructure as a Service (IaaS). Two real-life datasets were collected from 30 and 60 users for SaaS and IaaS studies, respectively. A thorough design and investigation of the biometric techniques was undertaken, including description statistics analysis and pattern classification optimisation. A number of factors were analysed to evaluate the impact on system performance, such as volume of data and

type of sample selection. On average, using random sampling, the best experimental result achieved an EER (Equal Error Rate) of as low as 5.8%; six users experienced EERs equal to or less than 0.3%. Moreover, the IaaS study achieved a higher performance than the SaaS study with an overall EER of 0.32%.

Based on the intensive analysis of the experimental performance of SaaS and IaaS studies, it has been identified that changes in user behaviour over time can negatively affect the performance of the suggested technique. Therefore, a dynamic template renewal procedure has been proposed as a novel solution to keep recent user behaviour updated in the current users' templates. The practical experimental result using the more realistic time-series sampling methodology has shown the validity of the proposed solution with higher accuracy of 5.77 % EER.

Contents

List of Figures

List of Tables

1. Introduction and Overview

1.1 Introduction

Cloud computing technologies have changed the delivery of IT resources into virtual services that are accessible through the internet using web browsers. Using these cloud services, customers can build and run projects, browse and buy products, send and receive emails, store confidential information, transfer money, communicate with friends, and watch videos. This gives customers the flexibility, efficiency, cost effectiveness, easy deployment and on-demand services they want (Mell and Grance, 2011; Prasanth et al., 2015). The 'pay-as-you-go' concept 'makes the cloud an essential technology of modern IT that provides an economic solution for customers and organisations (Florentine, 2016). As a result, many companies such as Netflix, eBay, Xerox, Etsy, and Apple have decided to shift their products into cloud computing by renting resources from Cloud Service Providers (CSP) (SmartDataCollective, 2013; eBay inc, 2010).

According to the National Grid, the UK's gas and electricity network has announced plans to move its own internal data warehouses to cloud storage (Danny, 2013). Moreover, According to the Cisco Global Cloud Index, by 2019, more than 80% of all data centre traffic will be cloud traffic and around 86% of all amount of processing will be achieved in cloud infrastructure services (Cisco, 2018). Additionally, cloud-based spending is predicted to be more than half of IT spending by 2022 (Space Data Centres, 2017).

There is no doubt that the flexible and convenient facilities of cloud computing services have changed our daily lives (whether people are aware of it or not); however, the biggest barrier that hinders the development and widespread use of cloud computing services are security issues. Security issues cause challenges both commercially and technologically. Although many security mechanisms have been developed to reduce security-related risks (e.g., hacking), service providers and customers are still concerned about cybercrime on cloud services. Hackers have used various techniques to gain access to victims' systems, thereby bypassing the systems' security mechanisms (Chou, 2013). This has been clearly demonstrated by many incidents that have targeted popular cloud computing service providers. Some are listed below:

- The Microsoft Azure cloud computing platform faced serious security incidents in March 2009, which led to a massive collapse and outage of the service for 22 hours, with a loss of 45% of user data (Chen and Zhao, 2012).

- Dropbox, one of the most popular cloud services providers, was hacked in July 2012; usernames and passwords of many users were stolen from third-party websites. These stolen credentials helped hackers to gain access to customers' accounts and misuse their data (Gupta et al., 2013).

- Apple iCloud was compromised in 2014 as more than 20,000 passwords of its customer accounts were stolen, which resulted in users' personal photographs, specifically those of celebrities, being leaked online (Cameron, 2014).

- Recently, the Code Space's Amazon AWS account was hacked and as a result, they have stolen the credentials of the company, moreover the attackers were able to access the information systems of the company and deleted the most data, backups, machine settings and even the backups hosted on the remote sites were partially damaged. As a result of these devastating attacks, the Amazon were unable to provide services to its clients (Cloud Security Alliance, 2016).

- According to Cloud Security Alliance, a number of security incidents occurred to a British telecom provider (TalkTalk) in 2014 and 2015, resulting in four million of their customers' personal information being disclosed (Cloud Security Alliance, 2016).

- Google's Gmail server faced attack in 2016; more than 272 million email addresses and passwords were stolen (Yadron, 2016).

It is clear from these incidents that cybercriminals can obtain access to sensitive information even with comprehensive security controls in place and dedicated security teams being allocated. A key issue is that the cloud services rely on simple authenticated login and remain accessible to users afterward for significant periods. Thus, arguably more intelligent security measures are required to support system security. Therefore, it is important to build strong security techniques to secure the cloud-based system from being compromised. To secure any system from being abused by unauthorised access to the system, a continuous identity verification system is needed to detect unauthorized access and protect the users' accounts from illegitimate use.

1.2 Research Aims and Objectives

The aim of this research is to create a continuous identity verification system for users of cloud services that operates beyond the point of entry and provides a basis for ensuring legitimate access of the system in a convenient and usable fashion. It will also importantly provide the service provider with an understanding of when services are being misused or abused.

To achieve this aim, the following research objectives are set:

- Design a series of experiments to explore the feasibility of deploying behavioural-based profiling on the top layer of cloud computing services (SaaS).
- Design a series of experiments for investigating the feasibility of deploying behaviour-based profiling on underlying layers of cloud computing services (IaaS).
- Analyse various practical and operational aspects of deploying a behavioural profiling system.
- Propose a continuous verification approach that can keep updating users' template dynamically in order to mitigate the effect of user behaviour change over time on the performance of the system.

1.3 Report Structure

Chapter 2 provides a background to the main concepts of cloud computing, such as the definition, common services, and deployment models. The chapter finishes with viewing the main popular cloud computing risks/threats.

Chapter 3 begins by reviewing biometric systems in terms of system components, requirements, techniques, and system performances. The chapter also discusses the processes behind biometrics, definitions of each trait, and influencing factors that might affect the performance of these biometrics, as well as evaluates the degree to which these biometrics can be used in continuous and transparent authentication.

Chapter 4 presents a comprehensive literature review on behaviour profiling with mobile phones, computers, networks, websites, and cloud computing services. The chapter also discusses the positive and negative factors affecting research methodologies and performance.

Chapter 5 introduces a number of experimental studies into the feasibility of behaviour profiling that have been conducted on users of the Dropbox cloud service, with the aim of identifying possible behavioural patterns that could be useful in user verification. Several pattern classification methods are applied, including statistical and artificial intelligence algorithms. A number of effective features towards success verification are investigated and the most appropriate classifier is identified. Moreover, a number of factors are applied to identify their impacts on system performance, such as volume of data and type of sample selection.

Chapter 6 presents an investigation into applying behavioural profiling in an IaaS-based infrastructure for misuse detection. To examine the feasibility of this approach within cloud infrastructure services, users' interactions with the cloud infrastructure application were collected. A series of experiments were conducted using supervised machine learning algorithms to examine the ability of detecting

abnormal usage. A number of factors that influence the performance of the machine learning algorithms were studied including the nature of classifier, volume of data, type of sample selection and required data to create users' templates.

Chapter 7 discusses a number of practical and operational aspects of deploying a behavioural profiling system. This includes analysing main issues such as users' behaviour changes that might face the model during a real-life adaption and suggesting a suitable solution.

Chapter 8 summarises the conclusions arising from the research and highlights the key achievements and limitations. It also contains a suggestion for future research and development of this research.

2. Cloud Computing

2.1 Introduction

In the past few years, cloud computing has become a new paradigm for hosting and delivering services over the internet. Users do not have to think about infrastructure, maintenance of resources, or managing issues. Customers can directly access the resources (hardware and software) of cloud computing services from anywhere and at any time without the need for specific knowledge about the resources via the internet. This chapter presents the definition of cloud computing, the characteristics of cloud computing with deployment models and service models, and, finally, the main security threats/concerns that are related to cloud computing.

2.1 Definition of Cloud Computing

The word “cloud” was first used by Google’s former CEO Eric Schmidt to describe the business model that was provided through the internet in 2006. Later, this word became popular, particularly as a marketing term to present different ideas of business (Zhang et al., 2010). According to Marston et al. (2011), Amazon started in 2006 as the first company to offer cloud services to its customers via Amazon Web Service (AWS). Other big companies, such as Google and Microsoft, followed Amazon by offering similar services (Fowler and Worthen 2009). In fact, the concept of cloud computing is not new; it is only a combination of various existing technologies (e.g., virtualisation, processing, distributing, and centralisation) into a network platform (Zhang et al., 2010). This combination

made cloud computing as a novel concept. Yet, there is still confusion about the actual standard definition of cloud computing even though several studies have focused on determining a universal definition for this technology. Vaquero et al.'s (2008) research is one example of these studies that compared more than twenty definitions of cloud computing with the aim of extracting a uniform definition. The study concluded that the proposed definition by the National Institute of Standard Technology NIST (2011) was the most widely accepted, which is as follows:

> *"a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction"*

Figure 2.1 below demonstrates the NIST's definition of cloud computing including five essential characteristics, four deployment models, and three service models.

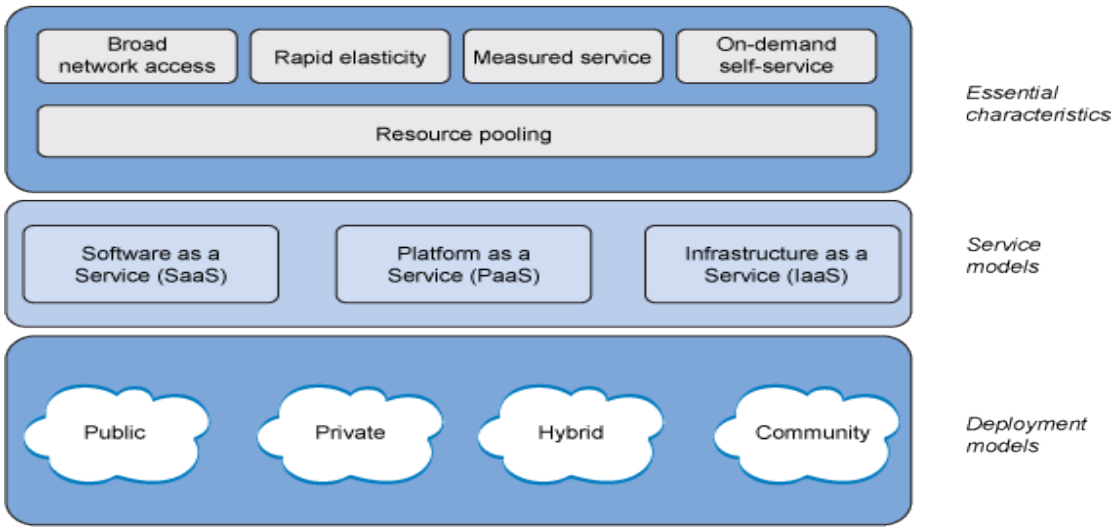


Figure 2.1: NIST Cloud Computing Definition Framework (Mell and Grance, 2011)

2.1.1 Characteristics of Cloud Computing

The main common characteristics of cloud computing, as defined by NIST (Mell and Grance, 2011), are:

- **Resource pooling:** cloud service providers offer sharing resource pooling that serves many users simultaneously. This technique is called the multi-tenant approach, where users can share the same service instance including different physical and virtual resources. They can assign and reassign dynamically according to their needs. The customers have no control or knowledge of where the resources are, but they may be able to determine a general location such as the region, country, or datacentre. Storage, memory, processing, network bandwidth, and application level are examples of these shared resources.

- **On-demand self-service:** costumers can request and use the provided resources by cloud service providers, such as server time and network storage, at any time without the need to interact with the provider.

- **Broad network access:** heterogeneous client platforms such as laptops, mobile phones, and PDAs can use the cloud services that are available over the network within standard capability mechanisms without being tied to a particular client.

- **Rapid elasticity:** the provisioned resources can be provided rapidly and elastically. Therefore, from the user's point of view, there is a sense that

the resources seem to be unlimited; customers can purchase any quantity at any time and quickly scale up and down for the resources.

- **Measured service:** cloud resources can be automatically controlled and optimised by cloud systems through monitoring and measuring the level of usage of these resources. This can help to provide a transparent report of the resources usage for both the provider, to support the availability, and the customer, to increase or decrease the amount of the resources.

2.1.2 Classification based on Service Models

Four deployment models have been classified according to NIST's definition of cloud computing, which are public cloud, private cloud, hybrid cloud, and community cloud, as illustrated in Figure 2.2.

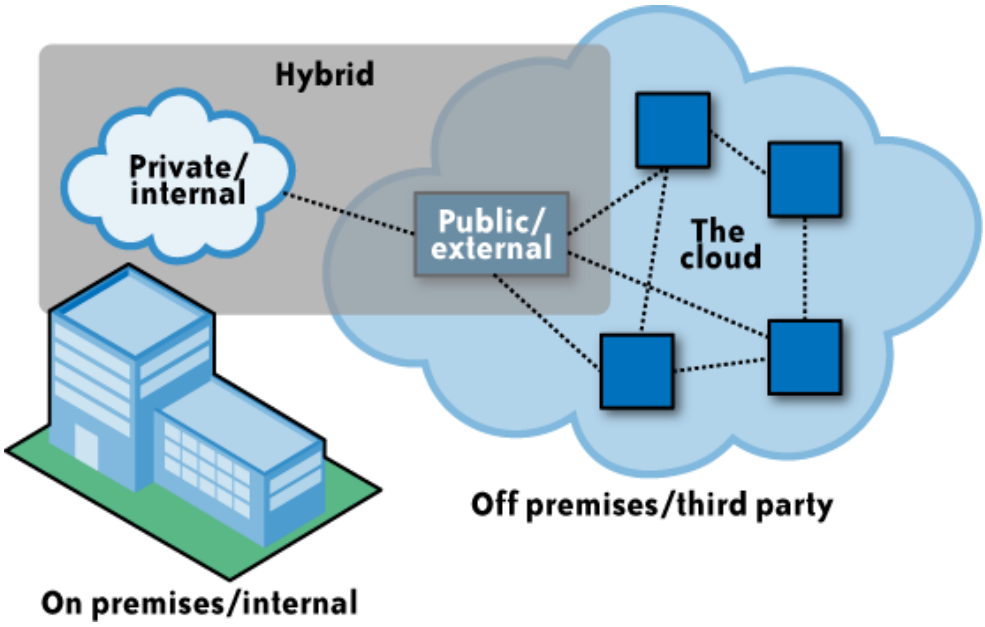


Figure 2.2: Cloud deployment models (Mather et al., 2009)

The providers of these models are different from each other through a number of aspects, e.g., some of them focus on lowering service cost whereas others are

interested in reliability and security aspects (Zhang et al., 2010). More details about these four models are as follows:

- **Public cloud:** also called external cloud, this model can be owned and managed by a third-party vendor (e.g., industry of a business, academic or government organisations). The providers provide the cloud infrastructure as a service for open use by the general users including small or large individual groups/industrial groups. Customers can interact with this service over the Internet through web applications that are offered by the providers. The main issue with this type of service is that the security management is done by the vendor. Consequently, customers have a lack of information about their data, including where it is stored, how the network is managed and what security measures are implemented. Elastic Cloud (EC2) by Amazon Web Services (AWS) is an example of the most popular commercial public cloud provider (AWS, 2018).

- **Private cloud:** also known as internal cloud, the infrastructure that is offered by this type is provisioned for a single organisation. For example, a large company may have this infrastructure to service its branches and customers. Thus, the offered service can be managed by the organisation itself, a special third party, or a combination of both (Mell and Grance, 2011). Although this model is costly to run compared to other models, it can offer the highest level of performance, reliability, and security for customers (Zhang et al., 2010).

- **Hybrid cloud:** this is a combination of public and private cloud computing models. As most services of this model can be deployed in the private cloud while the rest can be operated by public cloud, customers can benefit from the characteristics of both models. Nevertheless, the splitting of users' demands between the public and private clouds needs to be carefully determined, especially for peak demands. In this case, non-sensitive information can be processed in the public cloud while the critical application can be run by the private cloud through the careful management of peak users' demands. This known as "bursting cloud". These characteristics give the hybrid cloud more flexibility than the private and public clouds with more control of the application data. Additionally, this model can benefit from the scalability and cost-effectiveness of the public cloud by keeping sensitive data and critical applications under control.

- **Community cloud:** the cloud infrastructure of this model is shared by particular organisations of a specific community that have the same concern such as security, compliance, and policy (Mell and Grance, 2011). This type of cloud is typically owned and managed by single or group of members of the same community who are also responsible for the security requirements.

2.1.3 Classification based on Service Models

There are three main commercial cloud computing services according to the nature of the services are provided by cloud providers to their customers, as shown in Figure 2.3 (Mell and Grance, 2011).

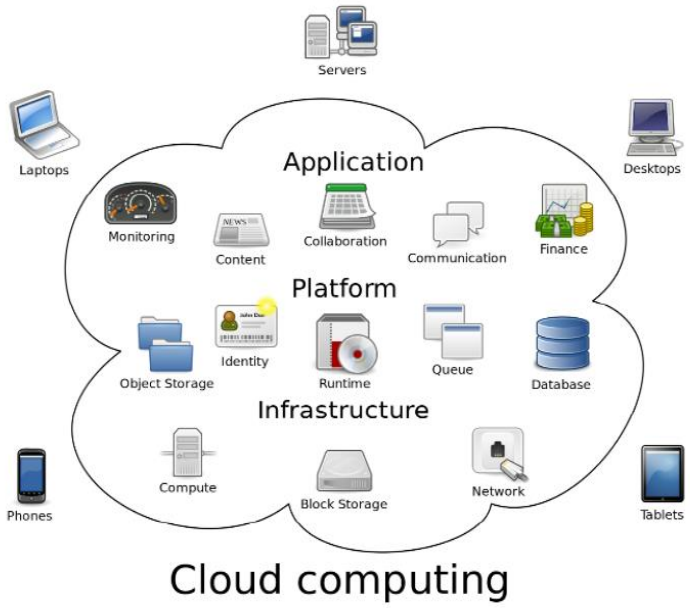


Figure 2.3: Cloud service models (Harikrishan, 2015)

- **Software as a Service (SaaS):** this business model is used to distribute diverse types of complete application software as a service according to customers' requirements. In this model, a single application is run remotely on the cloud server providers allowing many end users to use this service via the internet through different web browsers. Therefore, there is no need to create and build applications or software licenses as the applications are managed and maintained by the cloud service provider (Wu et al., 2009; Fernandes et al., 2014). SaaS can be used by ordinary users, as it does not require any technical skills or expertise (Jake, 2018). Google Applications, eBay, Dropbox, and Netflix are SaaS examples.

- **Platform as a Service (PaaS):** in this model, customers can use a complete development environment as a service, which is hosted on the servers of

CSP, without the need to install or configure this service. It allows customers to create and build their applications by using some programming languages. Customers can control their applications and may have some control over the development environment; however, they do not have control of the infrastructure layer, such as the operating system, hardware, or network (Eludiora et al., 2011). Aws Elastic Beanstalk, Google App Engine, and Force.com are examples of popular PaaS (Wu et al., 2009).

- **Infrastructure as a Service (IaaS):** this is a bottom-layer model; the CSP can offer virtual resources to their customers as a service in the shortest time by using virtualisation technology. The customers are enabled to manage and configure the infrastructure resources to their requirements with an on-demand and pay as you go service. These resources include the network devices, OS, storage, CPU, RAM and other computing resources. Moreover, customers can control the underlying infrastructure resources by allowing a cloud to build, run, stop, delete, or elastic virtual resources (Buyya et al., 2013; Fernandes et al., 2014). This service can be used by ordinary users and small/medium or large companies. Amazon EC2, Cisco Metapod, Microsoft Azure, and Google Compute Engine (GCE) are examples of IaaS.

In addition to the three core cloud sevice models, a variety of other cloud services exist to fulfil the needs of the companies, few of them are listed below:

1- Identity and Access Management (IAM) as a service: It refers to everything from password to user management and verification tools. This service is

hosted by a third part cloud vendor by providing a control to manage the user identity while interacting with the cloud services. One of the main advantage of this service is to provide Single Sign-On (SSO) functionality, which allows the users to login just once in IT infrastructure of company in order to accesses multiple services (Hemparuva et al., 2018).

2- Security as a service (SECaaS): This service can address the most of security issues for cloud computing services. This service deals with different levels of the available cloud-based applications to support a user-centric security approach. As a result of this service security mechanism of Cloud service providers is enhanced which gives more confidence to Cloud users in terms of securing their data in cloud (Hussain and Abdulsalam, 2011).

3- Network as a Service: this service allows cloud users to access the network infrastructure directly and securely to get additional resources collocated with switches and routers. This includes the provision of a virtual network service by cloud providers of this network to a third party and furthermore users can also choose the computational and storage resources that are needed for his/her applications (Shaukat et al., 2016).

4- Databased as a Service (DBaaS): It is one of the cloud computing services that offers cloud users to access a database without the need to set up on-premises resources. The manager component of DBaaS controls the all underlying database instances by an API. The user can access this API through a management console which can be used to manage and configure the database (Dinesh and Ramteke, 2015).

5- Recovery as a Service: It is also known as “Disaster recovery” as a service. This service is a category of cloud computing that can be used to protect users’ data or applications from natural or people disaster or any other disruption at one location. This service can offer full recovery in the cloud (Wood et al., 2010).

2.2 Cloud Computing Architecture

Cloud computer systems can be divided into two main sections, one on the front end and one on the back end. The front end is on the user’s side while the back end is where the system resources resides. These sections are connected with each other via the network over the internet. Users can see and deal with the front end to access the cloud resources of the back-end section, such as various computers, services, and data storage through browsers hosting on users’ computer desktops or mobile phones. There is a central server, namely middleware, which is responsible for administrating the system and observing the traffic by allowing networked computers to communicate with each other (Jadeja and Modi, 2012).

The cloud computing architecture can also be considered as a hierarchical layer that consists of four layers –the application layer, the platform layer, the infrastructure layer, and the hardware/data centre layer–as illustrated in Figure 2.4.

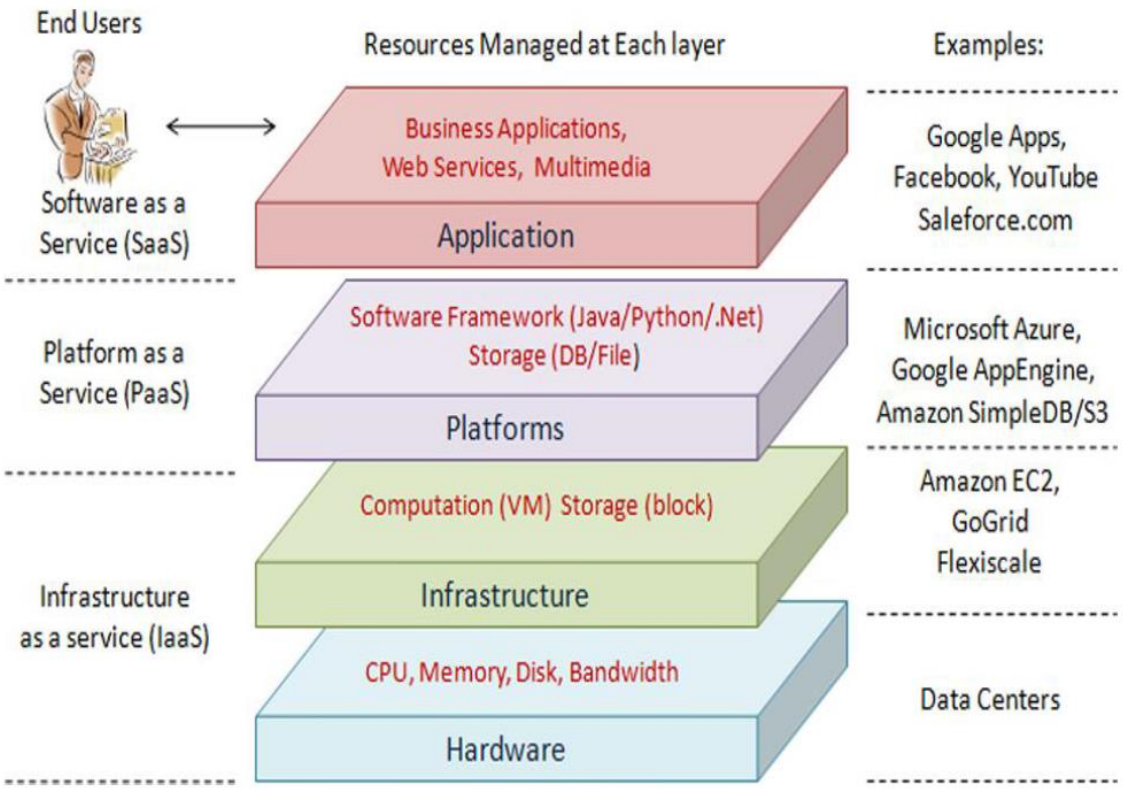


Figure 2.4: Cloud computing architecture (Zhang et al., 2010)

The underlying layer (hardware) that is implanted in data centres is responsible for managing the physical resources of the cloud (e.g., many servers are interconnected with each other via switches and routers). The virtualisation layer (infrastructure layer) is considered a large pool of computing resources that can be created and partitioned by a hypervisor such as Xen, VMware ESXI, Hyper-V and Oracle Virtual Box. This layer is an important layer in cloud computing architecture, as it handles the virtualisation technology such as dynamic resource assignment.

The third layer is the platform layer, which is located at the top of the virtualisation layer. This layer includes the operating systems and application frameworks. The

layer helps in reducing the burden of the direct deployment of the applications into the VM containers. Finally, the application layer is the top level of the hierarchy and can be used by users to leverage the automatic-scaling of cloud features, such as better performance, low cost, and availability.

2.3 Cloud Security Threats/Concerns

Cloud computing is becoming an attractive target for attacks owing to the distributed nature of the cloud (Khalil et al., 2014). Several studies have listed security as one of the biggest barriers to adopting cloud computing services (Zissis and Lekkas 2012; Hashem et al., 2015; Karame and Stavrou 2017; Gupta et al., 2017; Spanaki and Sklavos 2018). Cloud Security Alliance (CSA) has identified the top security threats that can be faced by any individual/organisation who uses cloud computing services (Cloud Security Alliance, 2017). The threats considered as the top cloud security threats for 2018 (Violina, 2018) are listed below.

1- **Advanced Persistent Threats** are unauthorised access to the cloud that aims to stay undetectable for a long time. The main target of this attack is to steal data through compromising infrastructures rather than to cause damage to the cloud services. Advanced persistent threats are automated pieces of code that can be inserted by coordinating with human involvement during interacting with cloud services (Chandra et al., 2014). These attacks are conducted through continuously monitoring and interacting with services which usually takes a long time to get a high value of digital assets that can bring a competitive advantage or strategic benefits such as government institutions, intellectual property, commercial secrets, etc.

The attacks may use different strategies, such as encryption, to obfuscate network traffic and zero day attacks to avoid signature-based detection– It refers to detect a specific type of known attacks such as known malicious instruction sequence used by malware that are predefined in datasets (Chen et al., 2016). Consequently, these types of attacks might be difficult to identify using conventional security measures, such as IDS and IPS (Intrusion Prevention System). One of the example of Advance Persistent Threats is Stuxnet which was part of one of the four highly complex malware that devastated the Iranian Nuclear Program. This malware caused a physical damage to the infrastructure without being detected for four years and because of this malware the efficacy of the nuclear plant was badly damaged (Virvilis and Gritzalis,2013).

2- **Insufficient Identity, Credential, and Access Management** is an issue that can allow illegitimate users to get permissions and access the cloud services. This attack can be used to download data, delete, read, and install pieces of software, and modify configurations. This happens owing to issues related to the control plane and management functions, which can cause potential damage to organisations or end users.

3- **Account Hijacking** is where an attacker's target is to hijack users' credentials and passwords for malicious purposes. Using this stolen personal information, attackers can access critical places in cloud computing services. Phishing and fraud are examples of this type of attack.

4- **Denial of Service** is a type of attack that uses botnet-triggered traffic to flood a website, preventing legitimate users to access their cloud services,

including data and applications. This type of threat can target specific users by changing their credentials or by entering many wrong passwords, which lock legitimate accounts or make them difficult to access. According to a report by the Hong Kong government, this type of attack compromised personal data of 77 million Sony customers in 2011, when they were denied their login rights (Hong Kong Government News, 2011). The effectiveness of this attack on cloud computing services gives cyber-attackers enough time to execute their operations without being identified. Due to the nature of cloud computing that can support multi tenants, Distributed Denial of Service (DDoS) attacks can be more devastating as compared to the on premise IT infrastructure. The reason behind this is that it can generate a number of flooding attacks simultaneously which results in unavailability of cloud computing services to the tenants (Yan et al., 2016).

5- **Data Breaches** are when sensitive, protected, or confidential data are released, viewed, stolen, or used by users who do not have the authorisation to access this data. Data breaches may occur because the cloud infrastructure lacks effective confidentiality measures. This can cause a negative impact on cloud computing services including loss of revenue, unexpected costs during responding to the act, and loss of customer confidence. Many cloud computing services had data breaches; for instance, more than 1.4 billion records were lost or stolen in March 2017 only (breach level index, 2017).

6- **Data Loss** is where data on the cloud is lost from the system for different reasons, such as being accidentally (or maliciously) deleted. This does not

have to result from a cyber-attack; it can arise through natural disasters, such as fires or earthquakes.

7- **Insecure Application Programming Interfaces**: as APIs act a public front door to the cloud computing services that users use to manage and interact with these services, a relatively weak or unsecure interface can give a cyber-attacker the ability to access the application of these services. Moreover, the availability of cloud services is dependent on the security of these interfaces, as many operations pass through them from authentication and access control to encryption and activity monitoring.

8- **Shared Technology Issues**: shared technology is applied by cloud service providers to deliver various services to customers, including sharing infrastructure, platforms, and applications. However, the flaws in a hypervisor—which has the responsibility of sharing technology—sometimes allows people to access the platform of other users because of weak isolation. This can enable attackers to access and use the shared memory and resources of legitimate users.

9- **Malicious Insider** this type of attack can be implemented by a user who has/had authorised access to the cloud services, such as a system administrator, employee, contractor, or another business partner. This user can exceed or misuse different aspects related to the service, such as network, system, or data. Base on the privilege of the user, he/she may attempt to access customer information by using their accounts without being identi-

fied. A study by Aleem and Sprott (2012) reported that 52.9% of ICT professionals involved in their survey said malicious insider attacks can be a major issue that can face the adoption of cloud services.

10- **Insufficient Due Diligence** often arises when an organisation rushes to adopt cloud technologies and selects cloud providers without performing due diligence. This can result in exposing the organisation to various risks including commercial, financial, and technical problems, thereby threatening its success.

11- **Abuse and Nefarious** often arises from poorly secured cloud service resources and cloud service trials. These resources can be utilised by malicious users through making fraudulent account sign-ups. Attackers can exploit these cloud features to undertake malicious activity across all cloud computing models (IaaS, PaaS, and SaaS).

12- **System Vulnerabilities** are exploitable bugs in software that hackers can use to steal data from the computer system. From the perspective of a cloud computing system, multi-tenancy technique for sharing the cloud computing resources can make the organisations' systems are placed close to each other. This can help attackers infiltrate and gain access to the shared memory and resources.

Thus, it can be noticed, based on the aforementioned threats, most of these threats are aimed at attacking any service that is connected with the internet and getting illegal access to the service's resources. Among the listed threats, the 2nd,

3rd, 7th, 9th and 11th threats can be implemented in any layer of the cloud computing models (SaaS, PaaS and IaaS). To use the cloud resources illegally, the attackers of these types of threats try to present themselves to the system as legitimate users by capturing users' identity information. Therefore, by stealing customers' login credentials, hackers can gain illicit access and misuse the service and user information. Additionally, according to a cloud security report by Crowd Research Partners in 2018, 91% of organisations are concerned about the security aspect of cloud computing services, particularly cyberattacks (Crowd, 2018). This type of attack is massive and data breaches have increased in 2018; the cost of a breach can reach up to 4 million US dollars. This causes enterprises to become unable to afford the cost of these types of attacks (Bennett, 2017).

It is clear from what is stated above that users' sensitive information within cloud computing services can be abused by cybercriminals even with security controls in place and dedicated security teams being allocated. The attacker can present himself as a legitimate user to the system through stealing legitimate users' identity information. Thus, customers would have concerns about unauthorised access to their information, which is remotely managed in these services.

2.4 Conclusion

This chapter showed the key attributes of cloud computing by presenting a definition of cloud computing, highlighting the combination of its characteristics, and explaining the main types of deployment and service models. However, owing to the distributed nature of the cloud, the system of cloud computing services is an easy target for cybercriminals, who are constantly looking for weaknesses in the

cloud system that can be exploited. This was clear from the listed threats and attacks that can violate and misuse users' cloud resources. Therefore, additional intelligent security techniques are arguably required to protect cloud services from being compromised and misused. A continuous identity verification system is a solution to protect cloud computing services by operating transparently to detect abnormal access. Users' biometrics can be used to monitor and evaluate the legitimacy of current users that interact with these services in a non-intrusive manner. Therefore, the next chapter will introduce all the aspects of users' biometrics, including physical and behavioural biometrics that can provide a better way to protect cloud computing services.

3. Biometric Systems

3.1 Introduction

Humans have used specifically unique characteristics, such as face and voice, to identify individuals for thousands of years; we can recognise friends and family through their faces or by hearing their voices over the phone (Prabhakar et al., 2003). However, the emergence of modern electronic technologies that are used to manage many daily tasks (e.g., accessing sensitive data, documents, and critical services) make it important to concentrate on the issue of personal identity. Therefore, biometrics have been applied in many modern technologies, such as smartphones, to build secure systems to avoid different types of attacks.

One of the approaches used to authenticate or recognise a person in many different security systems is the biometric approach. Biometric techniques can be subdivided into two types: physiological and behavioural biometrics (Gamboa and Fred, 2004). The physiological biometric recognition is based on the human body, such as the shape of a face, eye, or ear, whereas behavioural biometrics rely on a person's behaviours, i.e., the way they do a particular task, such as the way of writing their signature, walking, speaking, or other behavioural traits. In comparison with traditional knowledge-based (e.g., PIN) and token-based (e.g., bank smartcard) security systems, biometric techniques might be more complex to implement. Moreover, some external factors, such as the environment, accidents, and quality of equipment may affect the accuracy of biometric features. However, biometrics offer many unique features that might not be easily shared, stolen,

hacked, or broken by an attacker. As a result, biometrics have been used in many systems as an alternative security solution (Ross and Jain, 2004; Jain et al., 2007).

Biometric techniques can be used in continuous identity verification by monitoring the characteristics of the biometrics non-intrusively (Clarke, 2011). Non-intrusive verification (authentication) considers a biometric characteristic without users needing to explicitly interact with a system, which mitigates user inconvenience. This chapter will discuss the strengths, weaknesses, characteristics and performance affecting biometric systems, as well as the transparent nature of both physiological and behavioural biometric techniques, to provide insight into selecting the most appropriate technique to monitor and protect cloud computing services.

3.2 Biometric System Characteristics

Biometrics have several characteristics that can describe various aspects that can be used for different purposes, such as security systems. However, several factors should be considered when a particular biometric is used in a specific application. These factors can have a significant impact on biometric systems, such as matching decisions, level of uniqueness, and performance. Below is a list of the required characteristics for biometrics that can help to manage some of these concerns (Jain et al., 2007):

- Universality: this means every individual user who uses the technique should have the same trait. For example, if all users have fingers, it would be possible to use the fingerprint technique as a biometric identifier.

- Uniqueness: there are many unique traits of individuals. For example, the iris or retina is used for accessing military information because they are more unique than other biometrics, such as the face and finger.
- Permanence: is the ability to retain the biometric characteristics over time. Fingerprint attributes are an example of a physical biometric that remains unchangeable, whereas some behavioural biometric techniques are subject to change over time, e.g., the style of walking.
- Collectability: this attribute relies on the system and biometric type. Some systems need to process the collection of a biometric sample within a short or flexible time. Systems may also exhibit intrusive or non-intrusive patterns. For example, capturing the face of an individual can be achieved in a few seconds in a transparent mode using a normal camera while capturing an iris or retina sample requires a much longer time (and a deliberate action), which might be inconvenient (or acceptable) to some users.
- Performance: this refers to the accuracy, speed, and robustness of a biometric's performance a range of factors can affect the performance, such as the uniqueness and permanence.
- Acceptability: is a measure of the users' desire to present their biometric traits to the system. For example, some people may believe that retina or iris scanning may be harmful, whereas others may prefer fingerprinting.
- Circumvention: this considers the vulnerability of a biometric trait to being forged. For example, a retina scan is highly resistant to forgery or

spoofing while a fingerprint scan system can be tricked by using a fake finger (e.g. copy of the fingerprint).

It might be perceived that a biometric feature used in a biometric system should meet all the above requirements. However, human behaviour can be changeable over time because of several factors, such as mood, age, social environment, and health condition, which will affect the characteristics of the biometrics and the acceptance/perception of the underlying biometric method (Damopoulos, 2013). Therefore, great care needs to be taken with biometrics to select the most appropriate features that have most of the seven above-mentioned characteristics.

3.3 Biometric System

Biometrics have been used in many security systems to identify a person based on unique physiological (e.g., fingerprint) or behavioural (e.g., handwriting) features. As illustrated in Figure 3.1, there are five components for every biometric system (Clarke, 2011). They are as follows:

- Capture Component: This component considers the first stage of a biometric system, which includes capturing the biometric sample of a person using biometric sensor devices, such as a web camera, reading an individual's biometric features, and storing them as digital data.
- Feature Extraction Component: The extraction phase extracts a set of unique biometric features from the captured sample to generate a template and then stores this in a database. The unique features of the extraction stage depend on the system and biometric types, which can take several measurements for a particular sample.

- Storage Component: This stores the feature vector and other user information that will be used in the matching process of the authentication and identification systems.
- Classification or Matching Component: In this stage, the system will compare the new captured sample with the stored reference template(s). The output of this stage will show the degree of similarity between the two samples; the system will rely on their similarity to make a decision (accept or reject). This unit considers the main difference between the identification and verification mode. These two modes will be explained later.
- Decision Component: This is the final process of a biometric system; in this process the system will compare the value score of the matching stage with a threshold value of the system in order to make a decision. The result will determine whether a person will be accepted by the system. However, setting a threshold in practice is quite problematic because the system will rely on this threshold to make a decision. A poorly selected threshold will compromise system security by allowing an imposter to get access the system or denying the authorised user. Threshold value can be determined statically or dynamically. The static threshold can be applied when the same security threshold level is established for all transactions (Das, 2014). While the Dynamic threshold occurs when the security threshold is changed due to variances in both internal and external environment. It needs training or prior knowledge of the user interactions (samples) before setting the final value of the threshold. Therefore, an accurate threshold value can be obtained and this value can be also updated by retrain the

system after a period to make the system decision more accurate (Yan et al., 2016).

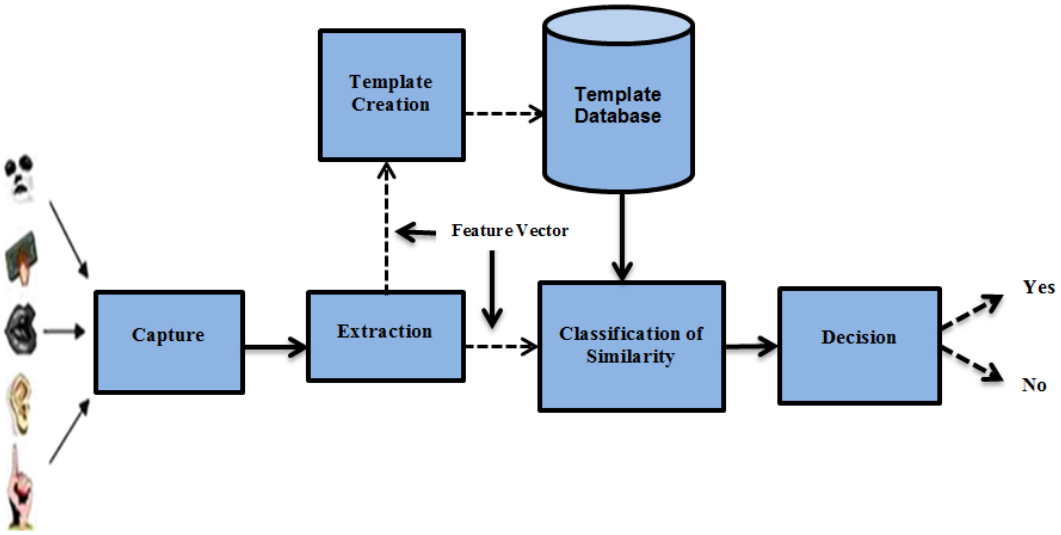


Figure 3.1: The biometric system process

Generally, these components will support the two process stages of a biometric system which are the enrolment and authentication/identification process. During the enrolment process, a person can register within a biometric system and the system collects the biometric sample from the person. This is conducted by an appropriate biometric sensor to create a reference template in a database. Subsequently, this template will be used in the matching process of the biometric system. At this stage, measures should be taken to ensure that only the authorised users are allowed to enrol and the samples are of good quality. This would improve the accuracy of the system and, hence, make sound authentication decisions (Jain et al., 2004).

After completing the enrolment process, the biometric system is ready to perform one of two distinct modes; the authentication/verification mode verifies a claimed identity and the identification mode determines the identity (Stephenson, 2009).

In the verification mode, the system matches the gathered biometric of a user with the claimed identity, which was previously stored in the template of the database system. If the matching is achieved, the person can access the system; otherwise, access is denied. This process of verification is called 1:1 matching. An example of the verification mode occurs when a person needs to login or access a computer system by using a username and fingerprint. First, he/she will type a username and then scan his/her finger. The recent capture sample of the fingerprint will be compared with the sample template of the previous fingerprint reference that was stored in the database based on the given username. If they match, the user can be granted access. Otherwise, the user will be rejected.

In the identification mode, all processes are similar to the verification mode, but the difference is in the matching process, where the user does not claim an identity but rather the system matches the sample against all enrolled users to identify if there is a match. For instance, if the police arrested someone and needed to find information about them, which had been stored in a police database, and access to this information was through his/her fingerprint, the matching process would compare the present fingerprint sample with all users' fingerprints in the database. This means the captured sample would be compared against every reference sample (one-to-many) to find the matching sample. Therefore, the

uniqueness of the features used for the identification system need to be more distinct than for the verification system.

Further, the identification mode usually needs a longer time than the verification mode because it involves more complexity and computation. More importantly, the identification system requires a higher level of uniqueness of biometric traits than the verification system to increase the system's accuracy. Consequently, behavioural profiling techniques are not recommended for identification systems.

3.4 Biometric System Performance

As previously mentioned, a biometric system can distinguish legitimate users from others based on the matching or comparison between the capture of the current user's biometric sample and the reference template sample(s), which are stored in the system's database. Several factors might affect the accuracy of biometric systems, such as environmental noise, which might lead to rejecting authorised users' access to the system.

There are two main error rates that can show the performance of biometric systems: the False Acceptance Rate (FAR) and False Rejection Rate (FRR), as shown in Figure 3.2.

- FAR refers to the likelihood that an imposter is falsely accepted by the system. The higher the FAR, the greater the possibility of an imposter entering the system. Assume that FA refers to the number of false accepts and NI refers to the number of imposter attempts, we can calculate the FAR from Equation 1.

$$FAR = FA/NI \quad (1)$$

- FRR represents the rate of the system rejecting legitimate users when they attempt to access the system. It is considered as an annoyance to authorised users because even though users have permission to access the system, they are denied by the system. Therefore, they may attempt to re-authenticate multiple times. Similar to the previous way, the FRR can be calculated by using Equation 2. Firstly, assuming that FR refers to the number of false rejects from the decision and NL is represented the number of legitimate user attempts, then the equation for FRR is:

$$FRR = FR/NL \quad (2)$$

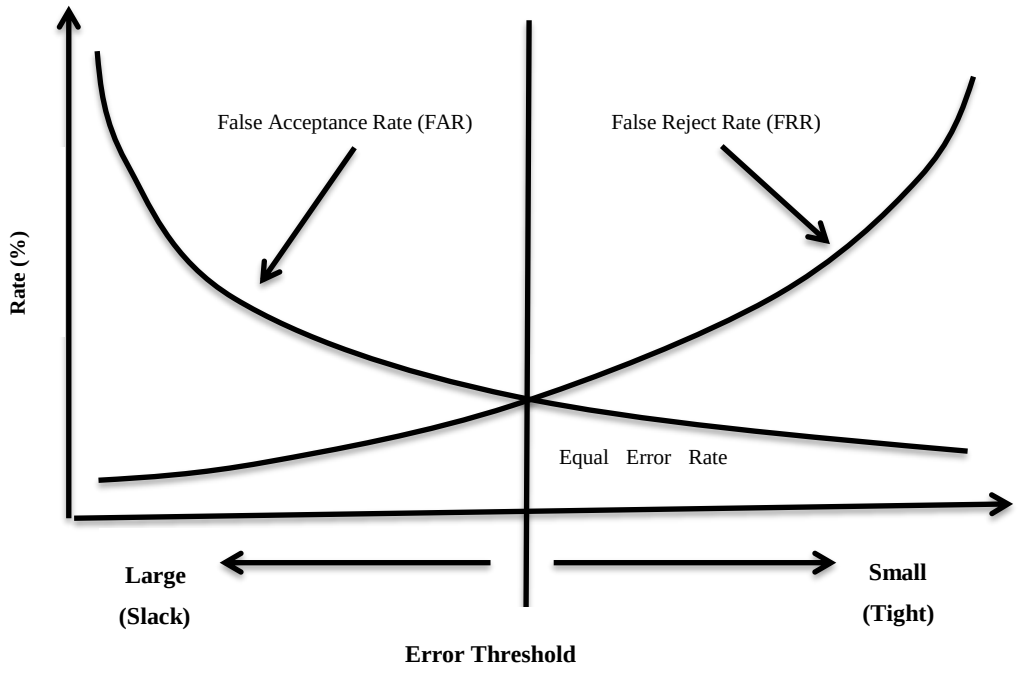


Figure 3.2: The Effect of Biometrics Performance Metrics Factors

Figure 3.2 shows the relationship between FAR and FRR is mutually exclusive because it is difficult to be both rejected and accepted simultaneously in the same authentication attempt (Adair et al., 2008). Dowland and Furnell (2002) reported that the relationship between FAR and FRR can be described as a trade-off relationship for system designers because when the threshold value is high (i.e., tight setting), the system security becomes strong but usability becomes an issue. When the threshold value is low (i.e., slack setting), the system security is reduced but the system becomes more convenient for the user. Therefore, with a low value of the FRR, the user will be granted access to the system with fewer attempts while, if the value of the FRR is increased, more attempts would be required for obtaining access. Generally, a threshold value setting that is required to meet security level and to be more convenient for the user is usually set (Clarke et al., 2002).

The Equal Error Rate (EER) is a third error rate and is defined as the point at which value the FAR and FRR curves meet or intersect; it is usually used to compare the performance between different biometric systems (Nanavati et al., 2002; Woodward et al., 2003). These performance rates of a biometric system are heavily based on the average of the test population: the larger the population set, the more reliable the performance. This means, with more users' samples used during the training, the more accurate result will be achieved. This is because a classifier learns more about users during the training stage and will also make an accurate decision when more samples are available during test the system. The FAR and FRR also rely on other considerations such as the uniqueness of each individual and the sophistication of the employed classification method (Li, 2012).

Generally, it is clear that if the value of the EER is low, the biometric system performance tends to be accurate.

On the other hand, the False Matching Rate (FMR) and False Non-Matching Rate (FNMR) are other biometrics performance metrics that can be used to measure the misclassification in the verification system of legitimate users and impostors. Although some literature (Jain et al., 2007; Woodward et al., 2003) has considered these metrics as synonymous with the previous metrics (FAR and FRR), they can be considered subsets of FAR and FRR, respectively (Clarke, 2011). This is because the FMR and FNMR can be used to measure the error rate within the matching or classification stage, whereas the FAR and FRR can be used to compute the error rate within the decision stage. Therefore, the FAR/FRR metrics are more inclusive in that they also encompass the Failure to Enrol rate (FTE) and the Failure to Acquire rate (FTA), which can be used to evaluate the biometric systems during the enrolment stage.

- The Failure to Enrol rate (FTE) refers to the failure rate during biometric registration of individuals in order to create the reference template sample (Jain et al., 2007).
- The Failure to Acquire rate (FTA) means the rate of unsuccessful extraction or capture of a biometric sample. This may be caused by the sensor of the device failing to capture the sample (Jain et al., 2007). When the FTE of a system is high, this leads to more effort on the user part during enrolment (Li, 2012). Various reasons may cause these types of errors.

For instance, accident effects (e.g., a missing finger), environmental effects (e.g., noise), ageing effects (e.g., changes in gait), changes in mood, and/or other factors. Therefore, these factors should be considered because the poor quality of the biometric sample capture will affect the accuracy of the system and its accessibility. This means the usage of the system might be inconvenient for their customers.

3.5 Authentication Methods in Cloud Computing

There are different authentication methods in cloud computing environment which are typically employed to improve the security of could services (Meena and Syal, 2017). These methods are:

1- Authentication via username and password: The primary goal of this method is to protect user data from unauthorised access, which is hosted in cloud. In this method of authentication, users must login with the provided credentials in order to access services hosted on the cloud.
2- Multi-factor authentication: In order to overcome the security weakness of aforementioned method, a multi-factor authentication method was introduced. This method not only rely on a user name and password, but also provide extra layer of security by taking in consideration other factors such as biometric authentication.
3- Trusted Computing Group: This method provides a group of properties based on hardware root of trust that has been developed by industry to protect computing infrastructure. It can offer a protection against viruses,

spam, phishing, physical theft and protect data, identity of users. An example of this service is Trusted Platform Module (TPM) and Mobile Trust Module (MTM). The TPM module refers to a security factor that can be adopted in PCs, whereas the MTM module is a security factor that can be used for mobile devices.

4- Public Key Infrastructure (PKI): Traditional security models were based on hiding a key by supporting traditional asymmetrical encryption method, such as RSA. Then it uses a private key to confirm the identity of legitimate user. PKI supports the distribution and identification of public encryption keys that can enable users and devices to secure exchanging the information over networks such as the internet and verify the identity of the other part.

5- Single Sign-On (SSO): It is an identity management system, which allows users to access to all services with only one time authentication token within an organisation. This means that users can access an independent multiple software systems within the same provider in single login without a need to re-login again to each system.

6- Biometric authentication: Physical and behavioural biometrics can be used to conf the identity of users. Physical biometrics such as face, finger, and hand were widely used in different applications for both verification and identification because their characteristics tend to be invariant and contain a high level of discriminatory features. This can make the distinguishing process more accurate than behavioural biometrics (Le, 2011). The behavioural biometric technique discriminates individuals based on the

unique patterns that are extracted from human behaviour, such as how a person types on a keyboard or the way of users interact with their services (behavioural profiling).

However, all of these aforementioned methods just focused on the point of entry; whereas they did not focused on the continuous monitoring users' interactions, while accessing the resources that is hosted on the cloud. Moreover, as the research focuses on the way of users interacting with application services to verify the identity of users continuously and transparently. In this research the behaviour profiling technique will be discussed in detail as this technique will fulfil the research requirements of this study.

3.6 Behavioural Profiling

Behavioural (or service utilisation) profiling identifies a person based on interaction patterns within a specific service or device, such as PCs or web applications. For instance, it can generate a behavioural profile for a person who uses a specific web application by determining the access time, duration, date, location, and sequence of events, as well as it could distinguish the type of applications through tracking the websites that are visited. Figure 3.3 shows the main behavioural profiling attributes that can be used to create user templates, such as time, date, and duration of using a website. The performance of extraction features to build the user's initial template is likely to be poor because users' patterns are difficult to obtain from little interaction. However, the user behavioural profile to verify a person possibly becomes strong over daily usage of the application, i.e., the period spent for system training to build reliable behavioural profiling for a user (Hogben,

2010). Therefore, more user interactions will help to build good discriminative patterns that can help distinguish among users.

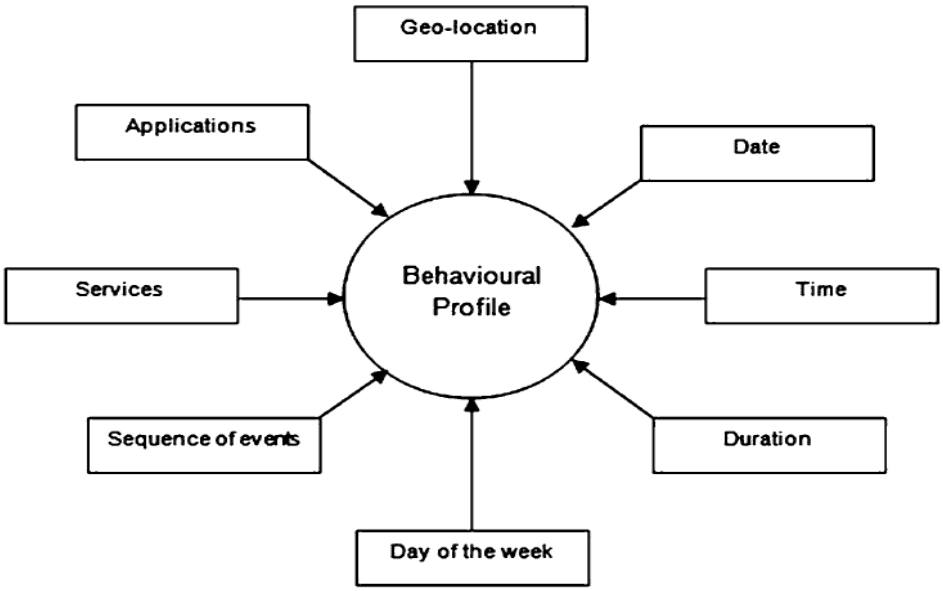


Figure 3.3: Behavioural profiling attributes (Clarke 2011)

Yang (2010) attempted to build a web user behavioural profile for user identification and mentioned that it can summarise much information about a user and store it in separated user profile. This information can be achieved and collected in two ways: explicitly and implicitly. Explicit information can be taken from users through the registration stages or surveys such as a user name, phone number, and address. This information can also be gathered explicitly through users' hobbies, such as the number of the user visits to some websites, the average spending on online purchases, and knowing favourite products. Implicit information is generated from analysing user activities via data mining or some popular statistical methods (Yang, 2010).

Yampolskiy (2008) reported that many different behavioural user profiles can be generated based on a specific software interaction to verify if the same user is interacting with this specific software environment or not. For example, “operating system interaction behaviour” can build a user profile to store some user behaviours when he/she starts choosing some tasks. Such as with Windows operating system, it can be concentrated on surveillance of a number of opened windows, the time between each window, and how many words are written in the window title. Another example of software interaction is “web browsing behaviour”, which can also build a great personal profile identifier by monitoring the set of actions during user interaction with the online web application, such as the selection of web browser type, keywords typing, and times of use (Yampolskiy, 2008).

Physiological biometrics can be used for both identification and verification while behavioural biometrics can be used for verification most of the time. Most of the behavioural biometrics can be applied for continuous and transparent authentication owing to their non-intrusive manner (Yampolskiy and Govindaraju, 2008). Various intrusion detection systems (IDSs) within Information Technology (IT) systems have used this technique by creating an alert to any abnormal behavioural usage during system use to protect the system from fraudsters (Yeung and Ding, 2003). Therefore, the behavioural profiling technique can be used to verify the legitimacy of current users transparently in a user-friendly manner.

3.7 Continuous and Transparent Verification

Traditionally, user verification systems are mostly based on the username and password to identify the identity of a person only in the point-of-entry phase; there

are no checks beyond the login stage of a service/session (Ceccarelli et al., 2013). The mere point-of-entry authentication means that the service remains open and available for the user during the entire working session. This can raise concerns about authorised access to user information within online applications at which cybercriminals have been increasingly targeting. Further, many impostors have attacked cloud computing services specifically. Many techniques have been developed to protect systems against different attacks; one of these techniques is to replace traditional knowledge-based approaches with biometric techniques (Kumar et al., 2005). Given that these techniques claim to be un-sharable and difficult to be stolen (besides other distinctive features), applying them may lead to improving the security level of systems. However, growing misuse and insider attacks, especially with critical applications such as financial and banking applications, have increased concerns. The problem is, when the system verifies the user, most of system resources become available to this user until the user logs out. In this case, if the user works with a group of employees in the company and leaves his/her computer opened and un-attended for 10 minutes, his/her computer can be easily misused by others and sensitive data can be accessed/stolen and un-authorised transactions can be carried out via his/her account. This issue could be managed by requesting that the user enter his/her credentials after a short period; however, this solution is likely to make the usage of the service inconvenient to the user.

Continuous identity verification techniques may offer a practical solution for the previous problem, knowing that users are likely to use these techniques owing to their user-friendliness (hence non-intrusiveness) (Crawford and Renaud, 2014).

A number of verification/authentication techniques require explicit interaction(s) of the user. Physiological biometrics is an example of these techniques, which tend to be intrusive, such as fingerprints. However, behavioural biometrics can be implemented for continuous and transparent identity verification, as they can identify a user without compromising on convenience, i.e., without prompting them to do any abnormal action specifically needed for the verification process. Moreover, it also does not need any knowledge-based information from the user, and does not typically need any additional or special hardware (Wang, 2010). Behavioural profiling is a good example of continuous and transparent verification, which has been implemented by a number of commercial companies to detect fraud on credit card and mobile calling systems. With these techniques, researchers have identified that the detection rates are more than 90% with low rates of false alarm, which may be up to 3% (Stormann, 1997; Clarke, 2011).

3.8 Conclusion

Biometrics have been applied in many security systems to identify and verify legitimate users from impostors. However, both of their types—physiological and behavioural—have strengths and weaknesses. Physical biometric methods offer high protection for systems because they have strongly unique biometric features that are difficult to change or forge and remain stable over time. In addition, they need less time to capture the initial and reference template sample. Therefore, they have been used in many authentication systems, particularly serving as a point of entry control; whereas the behavioural biometric approaches might be less accurate and unique and need longer time to build user templates. However,

they can be used as continuous identity authentication techniques to support and enhance the security system from misuse and insider attacks. Also, they tend to be more user-friendly; moreover, cost on hardware may not be required but other costs may exist; for instance, there might be a need to run software to extract users' features and generate templates, which might also need some additional storage to store these templates. Additionally, the system might need extra powerful CPUs to process these templates and reach a decision.

Additionally, as our problem is a cloud-based system to monitor the usage of the users continuously and transparently during interacting with cloud services to detect abnormal usage, most biometric techniques could not suitably be implemented for many reasons. For example, from a feature- capturing perspective, most physical biometrics such as the face, iris, and fingerprint are client-based. Therefore, it might be difficult to implement them with a cloud-based system because cloud providers might not have the authority to access the information of these traits. Moreover, physiological biometrics tend to be intrusive, such as fingerprints, which can make the usage of cloud services inconvenient to users. More importantly, additional cost, such as cameras and scan devices, might be needed to apply these techniques.

Further, compared to behavioural biometrics, such as keystroke, the behavioural profiling technique can gather much more information from users' interactions with cloud services through their web-based applications. Therefore, behavioural profiling can be considered as one of the behavioural biometric techniques that

can be used to extract many users' features through interaction with these applications to generate a user behavioural profile. As a result, it could be built to generate millions of user behaviour profiles for cloud computing service customers and track their behaviour usage. Subsequently, these profiles could help to improve and support the security systems of these services through continuous re-verification of the identity of these users without compromising on convenience. Therefore, the next chapter will discuss the main literature review of behavioural profiling techniques that have been applied for security purposes.

4 Literature Review of User Behavioural Profiling

4.1 Introduction

User behaviour profiling has been applied in various applications, such as IDS, fraud detection, and authentication—most of which are related to security. This technique is used to verify a user by tracking and storing the previous user activities, creating a user profile template(s) based on them, and then continuously tracking the user activities and comparing them with existing behavioural template(s) to make appropriate decisions on legitimate/illegitimate usage. This mechanism, therefore, seeks to increase the security level after log on by transparently re-authenticating users throughout the session. This would provide a user-friendly environment owing to continuous verification in the background with a non-intrusive manner during usage.

Although little literature was available on user-behaviour profiling in the context of continuous verification of cloud computing services, a considerable amount of literature has been published with other technologies, i.e., mobile phone systems, networks, computer systems, and web browsing. A critical evaluation of the literature of behaviour profiling is required to establish a greater degree of understanding of the domain. Therefore, this study evaluates the studies that have been done to better understand the most used techniques and methods; analyse the results that have been acquired and to know the main challenges and barriers that occurred with these systems. Ultimately, this literature review is sought to understand the potential feasibility of adopting behaviour profiling within a cloud environment.

The pivotal research question is how to verify a user while interacting with a service based on behavioural profiling. To identify relevant research, various keywords have been used, as in the following expression:

- (Fraud OR Intrusion OR Intruder OR Anomaly OR Misuse OR User Profile OR Behaviour Profiling OR Service Utilisation OR Application Usage OR Implicit OR Active OR Non-obvious OR Transparent OR Continuous) AND (Detection OR Authentication OR Verification OR Identification) AND (Mobile OR Mobile Phone OR Mobile Networks OR Mobile Device OR Telecommunication OR Smartphone OR Cell phone OR Computer OR Device OR Web browsing OR Website OR Web OR Cloud OR Cloud Computing OR Cloud Services).

This expression was used into the search engines of four well-known academic databases: IEEE, Springer, Elsevier, and Google Scholar. The earliest focusing studies of behavioural profiling were around 1997. A number of behavioural biometric techniques have been ignored such as voice, keystroke, and gait because of the limitations of applying their features extraction to the cloud environment. For example, keystroke analysis samples were captured on the client and thus it would be difficult to apply within a cloud environment. The outcome of this stage was 200 papers. By reading the abstract of these papers, it has been noticed that

many of these papers only shared similar keywords in their titles but the goal/purpose was different. Consequently, the final outcome of identifying the relevant studies was reduced from 200 to 46 papers.

4.2 User Behaviour Profiling for Mobile Phones

Around 1997, researchers started studying the possibility of applying user behaviour profiling to support and provide misuse monitoring for mobile networks. The earliest research on mobile phones focused mainly on IDS and fraud detection based on identifying the user behaviour activities during the interaction with mobile services, such as calling and mobility. Several recent studies have looked at other aspects, such as authentication to alleviate device misuse. The earliest studies are network-based, whereas more of the recent studies are host/device-based. An analysis of the work is presented in the following sub-sections.

4.2.1 User Behaviour Profiling of Calling Activity

The telephony service is considered as one of the main mobile phone activities and it contains a rich set of user features (e.g., start, end, duration of the call, local, international call, and dialled number) that can be used for fraud detection. These features can be investigated to build a user behaviour profile over time. If there is a high percentage of deviation between historical and current user-calling activities, it is a sign of potential misuse. Numerous studies have been implemented by using the calling activities to detect fraudulent attacks on mobile systems.

Burge and Shawe-Taylor (1997) and Moreau et al. (1997) can be considered the earliest project for detecting fraud in mobile communication networks, which were

a part of the ASPeCT (European Advanced Security for Personal Communications) project. The aim of ASPeCT was to solve future issues with telephony fraud for both Global System Mobile (GSM) and Universal Mobile Telecommunications System (UMTS) networks. The idea of these works was to generate statistical behaviour profiles to their users for detecting fraud. User behaviour profile creation relied on Toll Ticket to extract important information about calling activities (Toll Ticket is a bill issued, which encoded all calling activities of the mobile network customers), as illustrated in Table 4.1.

Table 4. 1: Feature Vector

International Mobile Subscriber Identity (IMSI)
Start date of call
Start time of call
Duration of call
Dialled calling numbers
Type of calling (national/international)

These selected features in the above table were recorded to create two types of user behaviour profiles, which were the current activities (short-term activities) and historical activities (long-term activities) of users' calls. By comparing these two profiles, the deviation ratio could be computed and then this value was compared with the threshold value of the system to make a decision.

Burge and Shawe-Taylor (1997) applied an unsupervised neural network method on the Vodafone mobile network in the UK to classify the call activity over two months; 75% of the fraudsters were detected while 4% of valid subscribers were

misclassified. Similarly, within the same context, Moreau et al. (1997) implemented a supervised neural network classification approach and the data was also extracted from toll tickets for 300 new users over a six-week period and 300 cases of fraud during a six-month period. The project detected 90% of the fraudsters correctly, whereas the misclassification ratio for rejecting legitimate users was 10%. The main problem of this type of data was the enormous irrelevant and noisy data that came from calling activities. This could affect the stage of creating the historical user behaviour profile during the training process. Therefore, when the system was tested, there was no guarantee of recognising legitimate users from fraudulent users because there was not enough information in the historical user profile to manage this issue. Consequently, it might be difficult to use this project in detection systems, particularly where a huge number of subscribers might lead to increased noise and, hence, errors.

Similarly, Samfat and Molva (1997) extended the two aforementioned studies and proposed “an Intrusion Detection Architecture for Mobile Networks” (IDAMN) applying a different approach for anomaly detection—a rule-based approach. The IDAMN involved tracking user behaviours of GSM mobile networks to track mobile impostors in terms of both calling and migration activities. The behavioural profile recorded the following information of users’ calling activities, as shown in Table 4.2. This calling activity is divided into two vectors (call and session). The call vector includes a local parameter that deals with collecting all outgoing calls, whereas the session vector measures a global parameter for many activities such

as the total number of calls, duration, handovers, and the duration of the network connection.

Table 4.2: User Calling Activity

Daily Usage	Calls per Week	National Call	International Call	Average Duration	Call Time	Destination Call	Origin of Call	Type of Call

To apply different intrusion detection algorithms, four user types were tested in the IDAMN by using a simulation method: domestic, business, corporate, and roamer to generate data from various mobile stations. A software was used to collect the dataset of this study; for each mobile user within a specific type of subscriber, a 300 intrusive session vector (each session representing one day of connection to the network) and 600 to 2000 call vectors. Table 4.3 below shows the false alarm and detection rates of call and session vector for the four user categories.

Table 4.3: False Alarm and Detection rates (Samfat and Molva, 1997)

Category of user	Call false alarm rate	Call detection rate	Session false alarm rate	Session detection rate
DOMESTIC	1%	67 to 100%	2%	80 to 100%
BUSINESS	1%	88 to 100%	2%	90 to 100%
CORPORATE	1%	60 to 100%	5%	87 to 100%
ROAMER	2%	82 to 100%	1%	95 to 100%

The above table clearly demonstrates that the session vectors had a better detection rate than call vectors did. The main reason for that is the session based method is over a long period and the deviation of an imposter's activities become more obvious than the behaviour of the legitimate user. Therefore, the business

category performance was the best among others with an average 89% detection rate and 1.5% false alarm rate.

However, Stormann (1997) used the combination of three techniques, including the supervised, unsupervised, and rule-based techniques, which are implemented in the previous works to evaluate the performance of each approach. The author applied an identical method using real data and the result of this evolution was that the worst performance was obtained by using the unsupervised classification method, which could be affected by the lack of training data. The performance of rule-based and supervised methods was better; however, a high level of fault alarms were also observed (as shown in Table 4.4).

Table 4.4: ASPeCT Performance comparison of classification methods (Stormann, 1997)

Classification Approach	Detection Rate (%)	False Alarm Rate (%)
Supervised	90 (60)	3 (0.3)
Rule-Based	99 (84)	24 (0.1)
Unsupervised	64	5

*Brackets indicate more practical performance

The study concluded that each classification approach contains some strong features that might be better than other approaches; therefore, the combination of these three classification approaches into a hybrid tool might provide a more robust system because the strengths of each method can handle the weakness of other approaches.

Additionally, Grosser et al. (2005) also worked on fraud detection in mobile calls by applying a Self-Organising Map (SOM) neural network to generate resembling group patterns of user behaviour profiles. The mobile call activities that were used

to build these profiles were: Local (LOC), National (NAT), and International (INT) calls. These profiles are used to detect fraud based on detecting unusual behaviour calls. Finally, the authors concluded that the users who were not utilising their mobile phones in the usual manner could be detected as fraud. However, it might be difficult to distinguish easily between abnormal and fraudulent calls. Becker et al. (2010) also confirm the previous conclusion of Grosser et al.'s work that the most challenging issue for fraud detection is recognising the fraudulent behaviour from unusual behaviour. It is, however, noteworthy that both studies had no performance rates.

Continuing with the same subject of fraud detection in telecommunication networks, Hilas and Sahalos (2005) attempted to involve more features of calling activities in their work and applied a statistical machine learning approach to build user behaviour profiles. The research included eight elements as an input vector for constructing user profiles. These were:

- Calls made to local destinations (loc)
- Duration of local calls (locd)
- Number of calls to mobile phone destinations (mob)
- Duration of mobile calls (mobd)
- Number of calls to national (nat)
- International (int) destination and their corresponding durations (natd, intd)

The current input vector was compared with the historical vector stored in the user profile to realise the similarity value. The authors suggested that the similarity of users' calls can be achieved at two levels. The first level is to examine the equality of the number of calls in the similar group while the second level is to compare the total call duration per each group. More than five thousand members of university staff participated in this experiment over one year and 22,000 phone calls

were collected. The highest accuracy of the similarity found was 80%. However, the study did not involve malicious behaviours because no fraudulent data was presented of all data that were implemented in the experiment.

In 2007, the same authors tried to improve the previous result by applying different methods and datasets and increasing the number of features. Hilas and Sahalos (2007) applied decision trees to find system thresholds that can be used to determine the diversion ratio between the normal and fraudulent usage in the telecommunication system of a large company. A real dataset was used in this experiment, which collected daily and weekly usage over an eight-year period, as shown in Table 4.5 and Table 3.6. Moreover, this dataset consists of legitimate and fraudulent activities; 107,050 call records from 300 users were collected. Seventy-five of these users were established as fraudulent users. The authors tried to separate fraudulent from legitimate use based on critical values (thresholds). The user behaviour with weekly usage was classified correctly at 85%, whereas daily usage was 65%. The authors concluded that the combined user behaviour gives better accuracy results in fraud identification. However, there is not comparison result that can show the combination of user behaviours can improve the accuracy of fraud detection.

Table 4.5: Daily usage of user behaviour calls (Hilas et al. 2007)

Calls	Duration	Units	MaxDur	MaxUnits

Table 4.6 : Weekly usage of user behaviour calls (Hilas et al. 2007)

Mean(calls)	Std(calls)	Mean(dur)	Std(dur)	Max(calls)	Max(dur)	Max(cost)

Later, the same authors tried to implement a third method (Genetic Programing method) on the previous work to improve and enhance the result of distinguishing between legitimate and fraudulent usage (Hilas et al., 2014). The same dataset of the above work was applied. The outcome of their work compared to the 2007 study also indicates the findings are similar with no overall improvement. Applying different methods to the same problems and features might give better processing speed, but the result remains similar. Therefore, it is evident there is a need to select appropriate features, which would lead to improving the system accuracy.

In a similar vein and continuing with calling patterns to detect fraud in mobile phones, Ogwueleka (2009) proposed a SOM neural networks and probabilistic methods to learn the call behaviour patterns of users. The author focussed on the idea of differential and absolute analysis approaches. The differential approach is based on monitoring any sudden changes in the current user behaviour against historical behaviour; whereas the differential analysis methods are based on distinguishing the usual usage of behaviour fraud patterns. The dataset of fraud calls was collected from 180 users over a 75-day period; whereas the legitimate call data was over 38 days. The experiment resulted in a 3% of false positive rate.

Further studies within the similar context of mobile fraud, Qayyum et al. (2010) proposed the change in customers' calling behaviour to detect fraud in mobile networks by generating two statistical behaviour profiles (current and historical profiles) for each user. The study involved two considerations while capturing the fraudulent behaviour; firstly, the fraudulent behaviour could be captured if there

was a deviation from normal calling, whereas the second idea was the user's calling behaviour might change because of life-changing reasons, such as changing house or moving to another city. The authors developed a neural network with multi-input layers with suggestion of using two neurons as output layers in place of one. The idea of using the two output layers was to detect both fraud and non-fraud with different threshold values, and the error rate would be different in both output neurons during the test stage. The experimental result of correct prediction was 70%, which is the best in the case of using five hidden neurons rather than using three, four, six, or seven hidden neurons. However, the idea of using the two output layers was lost because only a single layer served both the two output layers because the weights of the neurons were not completely isolated.

Further on the implementation of behavioural profiling, Hebah (2011) presented a new approach to detecting fraud in the telecommunication field based on combining statistical and rule-based approaches in an unsupervised manner. Two types of behaviour profile are considered to build and update user's behaviour: characteristics and interaction preferences. The user's calling behaviour characteristics such as caller and receiver number, date, time, duration, and max cost of call, whereas interaction preferences profiling, such as the data of new services, should be collected manually. In other words, some users might prefer stopping their services while others may prefer sending only messages for the alert purpose. The main aims of the proposal are to make user profiling more accurate, more adaptable, and to use less process time. However, the proposal was not tested to prove these aims in reality.

Recently, Subudhi and Panigrahi (2015) introduced another new approach to detecting fraud in mobile telecommunication networks by using One-Class SVM (OC-SVM) formulated in Quarter SphereSVM (QS-SVM) in an unsupervised learning pattern. The work focuses on implementing more user behaviour features (i.e., user-id, date, time, duration, type, location, and call frequency) to build a user profile and apply the concept of QS-SVM to improve the classifier. Fraudulent calls were discriminated from the normal behaviour of users' calls by training on the SVM. The idea is if any current user behaviour calls do not match historical normal calls, an anomaly is identified. The reality mining dataset is used in the practical experiment to test the proposed system and to compare the results of the two classifiers. The results illustrate the QS-SVM has a better performance than the normal SVM for the fraud detection system. The accuracy and true positive rate of QS-SVM are better (both more than 97%), and there is a smaller false positive rate (less than 6%) and less execution time (nearly one minute).

4.2.2 User Behaviour Profiling using Location

The information of user commuting and relocation is a vital feature for generating a user profile in mobile devices. Most of the earliest research depended on mobile network companies to provide this information on users' migration through travelling from one cell tower to another. However, currently, smartphones have highly accurate built-in GPS sensors, which can be used to record users' mobility. Therefore, a number of studies have investigated utilising this activity to generate user-tracking profiles, which will be used for reducing security issues in mobile phones based onIDS systems.

The IDAMN also monitors users' travel within the network to detect imposters; it can track users' migration through their movement from one cell tower to another and store the information in the user's profile. When these movement activities exceed the threshold of historical mobility information, it can be considered as an abnormal activity. 400 simulated users were tested; each user had 300 abnormal itineraries that were applied in their test. The results of this system showed that the best system performance was the domestic user type with a 2% of the false alarm and 90% of the detection rate. In contrast, the worst performance of the system was the roamer user category with a false alarm rate and detection rate 7% and 65%, respectively (Samfat and Molva, 1997). This is because users might change their movements, which can affect the performance of the system because the system did not include a dynamic learning process, leading to confuse the classifier.

Buschkes et al. (1998) proposed a Bayes decision rule method for an anomaly detection system of GSM mobile networks to increase the security level. This method relied on collecting user's mobility patterns to generate a user behaviour profile. The user behaviour was based on measuring the average of mean residence times of staying mobile within a one cell and also when entering another cell. This method can classify users into many classes depending on a few observable characteristics. The experiment was applied in two scenarios: a town and a motorway. The results of the two scenarios showed that the predication rate of the motorway scenarios was better (with more than 94%) than the town scenario (with more than 80%) within five days, and 95% and 83% after 15 days,

respectively. The system did not include a dynamic learning process that considers the actual user positions are already verified. Consequently, the system may not involve all the users, but it can be implemented for specific security requirements.

Further, Sun et al. (2004) proposed an online anomaly detection method to identify a special group of the internal attacks within the mobile networks based on a mobility pattern. Three combined techniques were used in their study. The first technique was a high order Markova model, which was used to calculate a mobile user's movements from one place to other. Secondly, a Ziv-Lempel data comparison algorithm was applied to build user profiles for the mobility and route-related information. Additionally, Exponentially Weighted Moving Model (EWMA) was implemented for updating the user profiles. The system can automatically compare the current activities of the user's mobility with the historical activities. if the diversion value exceeds a threshold system value, an alert can be generated. The simulation environment was within 40 cells; each of them has six neighbours on average and the distance between two towers of cells is one mile as an average. The simulation results of the system showed that by increasing the user's mobility, the false alarm rate will decrease and the detection rate will increase (as shown in Figure 4.1) because when the mobile user traverses many cells, the proportion of discrimination from others will increase. Moreover, the discrimination between the legitimate user and impostor will be increased remarkably at a high speed because the mobile will cover more cell towers.

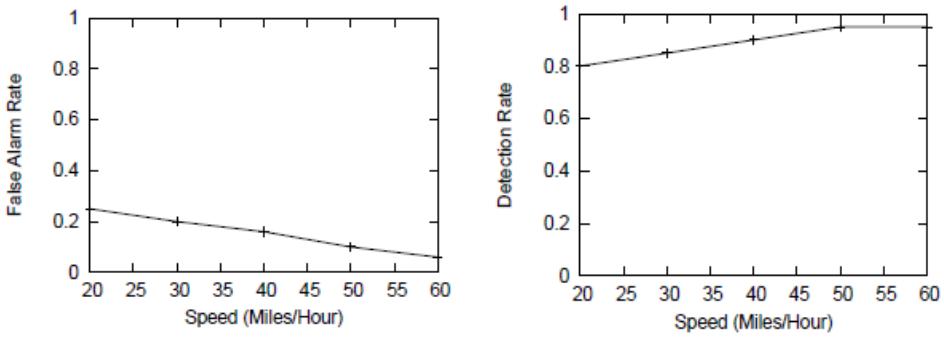


Figure 4.1: False Alarm Rate and Detection Rate at Different Mobility Levels (Sun et al. 2004)

However, the research did not consider average speed that is less than 20 miles per hour. This is clear from the previous figure that the false alarm rate and detection rate of the system will be poor; this means also the pedestrians or traffic jams and speed limits in the city were not considered. Further, the system assumed that the users' places and itineraries during his/her mobility are constant. These overlooked factors may expose some particular issues; for instance, a taxi driver cannot be detected easily by the system because they may pass through the varied directions and places daily.

Sun et al. (2006) tried to improve the performance of the previous system by adding several modifications. The new system modified the threshold value frequently when the set of movement patterns changed and two behaviour profiles were built for each user (weekday and weekend route profiles). The new results of the simulation method showed that on average, the false positive rate was 7%, which was less than their previous test, whereas the detection rate had a roughly similar performance. However, the performance metric for both the false positive rate and the detection rate remained the same for users who have a speed less than 20 miles per hour.

In a similar vein and continuing with mobility patterns to detect fraud in mobile phones, but with other methods and techniques, Hall et al. (2005) proposed an instance-based learning approach that benefited from public transportation for building users' mobility profiles to detect impostors. Fifty mobile users of public transport in Los Angeles were tested and evaluated by using a simulation approach and the result was 50% for the detection rate and false alarm. This migration is called the migration itinerary-based mobile IDS system, which monitors all cells that users have covered during their journey from one place to another. However, the performance shows the accuracy of the system is low and the mobile user who does not use public transport cannot be managed by the system.

Yazji et al. (2011) tried to detect anomalies in mobile phones for protecting data saved on mobile devices by combining spatiotemporal information and trajectory analysis. The spatiotemporal information includes extracting and building user profiles based on (1) user ID; (2) location information; and (3) timestamps, whereas the trajectory analysis considers time-location sequences of recently visited places. This combination is to build the relationship between location and time of day. Based on the reality mining dataset, which covers data for more than 100 users over a nine-month period, the accurate detection of the system was 81% with a 15-minute delay time. In 2014, the same authors tried to extend their previous work to acquire a better result. They applied two statistical methods to generate normal user behaviour profiles. The first method depended on the measurement of a practical cumulative probability and the second method used Marko

properties of trajectories. The accuracy of system detection was better than previous work, which was 94% within 15 minutes (Yazji et al., 2014). However, the study did not involve the capturing of visiting new places.

Most previous studies (calling and mobility activity) tried to improve the system accuracy detection based on mobile networks; the results were approximately similar. Many other facilities have been developed than call and text last 10-20 years on mobile devices, which can collect more information from the device directly. Several recent studies have investigated these facilities to increase the accuracy of detection systems, as demonstrated in the next section.

4.2.3 User Behaviour Profiling using Application Usage

In the last few years, many mobile applications have been developed. Some of these applications are standard while others need to be installed on the device. This means, much more information can be gathered from user activities while interacting with these applications (e.g., phone calls, GPS locations, SMSs, emails, website visits, and calendar activities). These activities can be exploited to build an accurate behaviour profile, which can be investigated to increase the accuracy level of the security system for a device or application itself. A number of studies have mainly focused on continuous and transparent authentication aspects to reduce the risk of attacks on mobile phones.

Jakobsson et al. (2009) developed user behavioural patterns for implicit authentication to protect mobile devices. The core idea of their work was based on recording historical calls, SMSs, emails, calendar events, and GPS of the user's

locations and then calculating the differential score between the current and historical behaviour profiles of a user to obtain the deviation ratio. A weighted linear function was applied to find the differential score, which was ultimately compared to a predefined threshold score of the system to determine the authorised user. They concluded two important results; the first was that the authentication's security and usability can be improved based on behaviour profiling. Secondly, mobile devices have a rich set of data about user behaviour activities, which are suitable for implicit authentication. However, their study was not practically evaluated.

In a complementary vein, Shi et al. (2011) investigated more mobile activity features of user behaviour, which are calls, SMSs, historical website browsers, and GPS locations for continuous and transparent authentication to detect anomalies on smartphones. The scoring method to verify a user is dependent on identifying the good and bad events that can affect the system's decision. The good/positive events are those such as phone calls, sending and receiving text messages from regular phone numbers, or visiting familiar websites/locations. These good events will increase the degree of verification score; whereas the bad/negative events are those such as mobile calls/text messages from an unknown number and visiting unfamiliar websites/places. These events or behaviours will decrease the user verification score. The experimental evaluations involved 50 users to model genuine users over a two-week period; 60% of data usage was for training and the remaining data was for testing. The authors mentioned that combining multiple features is more powerful than depending on a single feature alone with fewer probabilities of false rejects. However, there is no practical result that shows or

proves what the authors mentioned about the performance. Moreover, all these activities were recorded by the device; therefore, the attacker could delete, change, or stop the recorded information.

Damopoulos et al. (2012) implemented and evaluated the performance of four machine learning methods (i.e., Bayesian network, RBF, KNN, and Random Forest) to detect anomalies service usage (misuse detection) in mobile phones. They collected the normal usage of the data logs from 35 iPhone users, which consist of 8,297 mobile calls, 11,321 SMSs, and 790 hyperlinks of web browsing usages. These behaviour features were recorded either individually (each application service is logged separately) or by merging all data of application services together and then using them for authentication purposes. The first highest accuracy algorithm of their work was the K-Nearest Neighbours with a true positive rate of 99.8% and an accuracy rate of 99.5%. The second highest accuracy result was given by Random Forest, which was 99.8% and 98.9% for a true positive rate and accuracy, respectively. However, compared with cloud data, the main limitation of their study was the dependence on specific operating systems (i.e., Windows and Macintosh OSs). Moreover, the strength of the used features set was not discussed.

Similarly, by providing authentication in a transparent manner using user behaviour profiles, Li et al. (2010) also tried to investigate the development of different mobile applications and services to build a user behaviour profile for authenticating misuse. The authors proposed an approach for an anomaly detection system, which was based on multi-level behaviour profiles by considering different mobile

applications: calling activity, device usage, and Bluetooth scanning. They used a radial basis neural network for learning user behaviour profiles and identifying impostors. In their study, the experimental findings of EER were 13.5%, 35.1% and 35.7% of telephony, device usage, and Bluetooth scanning, respectively. However, their system was tested on the MIT Reality data, which was collected from user activities on the Nokia 6600 mobile devices during September 2004 to June 2005. The dataset was old if compared with the new dataset of modern mobile phones in the same year of their work, for which many features can be extracted. The year after, by changing some application usage features, the experimental results for voice calls, SMSs, and general application usage were an EER of 5.4%, 2.2% and 13.5% respectively by using the same MIT dataset (Li et al., 2011).

Later, they developed their work by combining a rule-based method, dynamic profiling approach, and smoothing function to get a better result. An EER of 9.8% was achieved in their experiment by applying the same previous dataset (Li et al., 2014). However, the dataset that was used to evaluate the system was old, dating back to 2004, and contained a limited number of mobile applications.

Recently, in the same context of active authentication to detect impostors in mobile phones, Fridman et al. (2015) focused on Android mobile devices to generate user profiles from some application and service activities, such as application usage, web browsing, and GPS location. The Chair-Varshney optimal fusion rule was proposed in their work to asynchronously integrate selected features and to combine multimodal decisions of all the selected activities. The main differences

of their work to Li et al.'s (2014) were: using a novel modern dataset of the Samsung Galaxy smartphone, which contains behaviour features for 200 users during a 30-day period; secondly, the authentication accuracy performance of the system achieved an overall EER of 5% within one minute and 1% after half an hour of the user interactions with his/her smartphone. Moreover, the system accepts the addition of other classifiers to examine the overall improvement in the system performance without changing any rule in the decision fusion rules.

4.3 Client Side Behavioural Profiles

A number of researchers have focused on generating user behaviour profiles from device usage and file access activities of the computer system in the continuous and transparent authentication manner.

Aupy and Clarke (2005) presented an authentication technique to determine a unique behavioural profile for a user through natural interactions with the PC to provide a non-intrusive and continuous technique to verify the user. The user activities of their preliminary study were user interaction, which applications were used and when, and which websites were visited. The system used the 'logger' application to capture these user activities; as shown in Table 4.7.

Table 4.7: User events extracted by Logger

Feature	This action is continuously updated
KEY	When and where the word has been typed, the word is recorded.
OPN	The name and class of the window are recorded, when a window is opened
CLO	The name and class of the window are recorded, when a window is closed

A neural network classification known as the “feedforward multi-layered perceptron (FF-MLP)” was used in their project, which has a good classification performance to solve complex non-linear issues. The experiment result of the work was as an overall average 7.1% of EER; however, some issues were raised in the study such as, the small number of the participants (i.e.,10) and lack the richness required to obtain a reliable statistical classification result, which led to repeating the number of previous user interactions to manage this issue. Additionally, the study used authentication windows of 10 minutes, which might not be suitable because an impostor may be able to achieve misuse within a short time without affecting the authentication result. Therefore, system abuse may occur and the impostors’ actions will be considered as legitimate over time.

Yazji et al. (2009) proposed a new mechanism to identify a user implicitly by re-authentication of a user who works on portable computers. The study used K-means clustering to build user behaviour profiles from log data of access of the network and file system activities. The normal usage of user behaviour was built depending on the daily frequent access to the network and file system activities;and the authentication mechanism was deployed on the external service (server) not on the device itself. The proposed system’s accuracy was approximately 90% every five minutes of the time window. It can evaluate the system as a strong approach when the attacker does not know the frequent number of the user visiting websites, which had been considered as normal visiting. However, if imposters knew the pattern, they could change their behaviour to how the system behaves. Therefore, the system can be attacked easily by the attacker and the authenticated users who do not have regular daily access might be rejected from

the system as imposters. Another limitation of their system, as the authors mentioned, is that the training period was short (two weeks) and the FRR and FAR were high (11% and 13.7% respectively). Therefore, they suggested combining other features to make their authentication system better.

Furthermore, some of the studies investigate device usage and file access activity to detect the insider attacks. Salem and Stolfo (2011) mentioned that genuine users of the computer system have an idea about the location and structure of their files and how they search for specific files, so the way of searching is limited and targeted. Therefore, it can monitor how, when, and how much a user is accessing their files and information to build user behaviour profiles to detect any illegal access. However, some types of insider attacks are more difficult to identify because attackers are familiar with the structure of files. Sometimes, the insider attack happens when employees change their jobs and still have access to sensitive company data. Therefore, insider attacks are difficult to detect.

Hu et al. (2011) concentrated on identifying malicious data exfiltration activities of the insider attackers. Statistical methods were used to construct user profiles which were based on monitoring and analysing the valid usage of file repositories while accessing logs of legitimate users. Their experiment involved 23 random users who used software developer repositories for 30 days; by comparing the historical file access logs of these file repositories with recent user access activities, the deviation ratio of these activities can be identified. For example, a user's downloading activities can be monitored during a period of time and in the event

of a significant amount of these downloads exceeding the unusual previous pattern, it can be concluded that abnormal behaviour occurred. However, there is no result that shows the ability of this system to detect anomalous activities of insider attacks.

Continuing with the insider attacks on accessing files in the PC environment, Salem and Stolfo (2011) modelled user behaviour of searching for patterns and information access activities for the masquerade detection. A one-class support vector machine technique was used to model and train the users' searching behaviour. The authors monitored the average deviation between normal and abnormal behaviours of the user search. For the study, a dataset was collected from 18 users as normal user behaviour search activities and 40 users of a simulated masquerade data during four days; also each user was recorded with more than 500,000 records. The size of this dataset is 10 GB, is available publicly, and is called RUU (Are You You). This dataset was created by a Windows host sensor. The following activities were mainly logged, including all registry-based activity, the create and destroy process, GUI and file access, and DLL libraries. The scenario of masqueraded data collection was performed by asked the masquerader to do one of the following scenarios with co-workers:

- Malicious: masqueraders were asked to search in the friends' computers and find financial data within 15 minutes.
- Benign: masqueraders used the friends' computers as a legitimate user.
- Neutral: the masqueraders were left free to choose whether to access the friends' computer.

A single pre-configured machine was used to simulate these attacks in the lab. According to their assumption, the experiments demonstrated that this approach achieved 100% detection of the simulated masquerade attacks with a 1.1% false positive rate with a two-minute latency. The main advantage of using this technique is to preserve users' privacy because it has the ability to build a classifier without sharing data with other users. However, four days of collecting the dataset was too limited and the lack of explanation about the methodology's details raises concerns about the credibility of the results.

Stolfo et al. (2012) tried to extend the previous work by implementing decoy information to reduce the insider attack to achieve high accuracy; in other words, the authors proposed combining the behaviour profile technique with decoy information for verification. Their study suggested the user behaviour and decoy documents stored in the cloud can be used as a certain sensor to detect illegal access. The idea of the study is that when the system informs the cloud system there might be an attacker the cloud system can verify the suspect by using decoy challenge questions, which will help to further identify the insider attacks.

Recently, within the same context, other research also confirms the previous proposal of combining the behavioural profiling technique with decoy information as a fertile approach for managing the risk of the insider attacks (masquerader) in cloud computing services (Sudha et al. 2014; Kanna et al., 2015) .

Some of the activities and techniques proposed in Stolfo et al. (2012), Sudha et al. (2014), and Kanna et al.'s (2015) research could be applied to cloud computing

services. Nevertheless, the ability of implementing their dataset to detect insider attacks in cloud systems might be difficult because of the following:

- The dataset was collected from one platform (Windows OS), which needs to be changed whenever the platform is changed, whereas the cloud computing system has heterogeneous independent OSs and networks.
- Attackers' activities on cloud computing platforms are service independent.

There is no standard dataset available for testing masquerade attacks on cloud computing services (Alguliev and Abdullaeva, 2014). More studies and information about applying user behaviour profiling to secure cloud computing will be discussed in subsection 4.5.

Regarding accessing activities, Abu Bakar and Haron (2014) proposed an adaptive authentication approach based on analysing the user login of the computer system to generate normal behaviour profiles. Four factors are considered to build these behavioural profiles: time of login, user's geolocation, application that was accessed, and type of web browser/OS being used. The system tries to identify a high risk in the illegal login attempts by comparing the normal and historical behaviour profiles of a user login. Moreover, the highest and the lowest security level for the authentication system were also considered. For example, an online banking application would be considered as a high security level and given a highest trust score, whereas other applications may not need to be considered at the highest security level such as news websites. However, the study was merely a theoretical proposal without any practical experiment.

4.4 Server Side Behavioural Profiles

Several studies tried to investigate behaviour profiling to build a user identifier by using user web-surfing activities.

Several studies focused on merely generating user behaviour profiles to understand the “patterns of interests” of websites that were visited by users. Mushtaq et al. (2004) used web usage information to build non-obvious user profiles from numerous log files of websites. The idea of their work assumed that the website is static and the website’s vender defined a list of topics; these topics are linked to one or more web pages. User behaviour profiling was based on spending time on various topics on the website. The work was just on one website and the website might be changed significantly. Similar but more thorough work was conducted by Zhang and Shukla (2006). They generated web user profiling from the data usage of several web services using a rule-based platform. The characteristic of the platform was not related to any particular application. A sequence of events was considered through generating action by users on a web service at a specific time, such as collecting the frequency of visiting a news website or downloading a music file. However, the work also was dependent on static website content, not dynamic content, to build a user profile. In addition, it was not implemented in the security field.

However, other attempts implemented similar previous work features (e.g., site names, number of pages, starting time and duration time of sessions) to build the user behaviour profiling for the implicit authentication purpose. For example, Yang (2010) applied a user-centric web browsing activity for a group of known

users to extract some users' activities of a website that have been visited frequently to build user profiles. 50,000 volunteers were involved in the study over a one-year period as a dataset that was taken from a commercial data vendor. Three methods were applied to build a user-behaviour profile: the support-based profiling method, the lift-based profiling method, and the learning tree or support vector machine (SVM) approach. From experiment results, Yang (2010) concluded that using the support and lift approach could be more accurate to identify users than the tree or SVM approaches. However, the study did not involve all users in the practical experiment. Seven users only who had at least 300 sessions in the dataset were selected to test the system. Moreover, the maximum number of users utilised in the dataset of the experiments was 100, so it is difficult to use this system to solve large-scale problems.

In another attempt to exploit the advantageous features of websites in the transparent authentication technique, Abramson and Aha (2013) used behavioural profiling of web browsing in the continuous authentication technique through monitoring user activities on a website over time. Global and internal session features were used in their study. The global features include the day of the week, time of day distributions, and the total number and duration of page views, whereas internal features consist of spending time on a web page, the time between the first and second pause of page views, and the time of revisiting the webpage. A one-class SVM classifier was applied in each feature set for each user. The experiment's results for global features were 56.5% and 37.1% as an average for FRR and FAR, respectively, whereas the overall result of internal features was 52% of FRR and 40% for FAR. However, based on the results, the error

rates were high, which means the features that were selected might not be enough for distinguishing a user from others; in addition, only 10 users were used for their experimental study.

4.5 User Behaviour Profiling in Cloud Computing

Any user can directly access the resources (hardware and software) of cloud computing services over the internet without the need to have specific knowledge about the resources. However, trust is one of the most important aspects of cloud computing, which is considered the main obstacle faced by the development of cloud services. This trust between the tenants and cloud providers is still poor (Li-qin et al., 2010; Alhanahnah et al., 2018). Therefore, several research studies have attempted to improve this trust by increasing the security level of the cloud environment by considering the flexibility and transparency of access control without interrupting or disturbing the cloud user. User behaviour profiling is one of these techniques that have been applied for this purpose. Due to a lack of studies on applying user behaviour profiling in cloud computing, studies with anomaly behaviour based IDS can be discussed as user behaviour is a part of anomaly IDS (it is focuses on detecting any anomaly behaviour of applications, services, programs and users.

Li-qin et al., (2010) proposed the “divide and treat” approach, which divides user behaviour trust hierarchically into small sub-trust layers and then each of these layers is divided into smaller data sub-layers. As shown in Figure 4.2, user behaviour trust can be divided into four sub-trust, which are: contract behaviour sub-

trust, expense behaviour sub-trust, identity re-authentication sub-trust and security behaviour sub-trust. The contract behaviour component is to check whether user behaviour is within the contract that made with cloud provider or not. Cloud users should use the cloud resources according to cloud regulations. While the expense behaviour component is refer to check the cloud users whether are within the agreed term of resource consumption or not. The third component (identity re-authentication sub-trust) is refer to re-authenticate the cloud user to prove the identity. This can be happen when a user lost his/her devices and the user name and password set up as default. Therefore, cloud providers can re-authenticate the user when they notice any abnormal behaviour. Whereas, the security behaviour sub-trust is to refer to whether user behaviour tries to attack the cloud resources. For example, Denial of Service (DoS) can be sent in the name of the authorised user. The value of user behaviour trust ultimately will be collected by using statistical analysis from aggregating the results of each level of the layers, as shown in Figure 4.2. After computing and evaluating behaviour trust strategy for each access in cloud services, the server will send the value of evaluation to stations for making an access decision. However, the study is only a theoretical proposal without any practical experiment being applied to prove the ability of this technique to achieve their proposal.

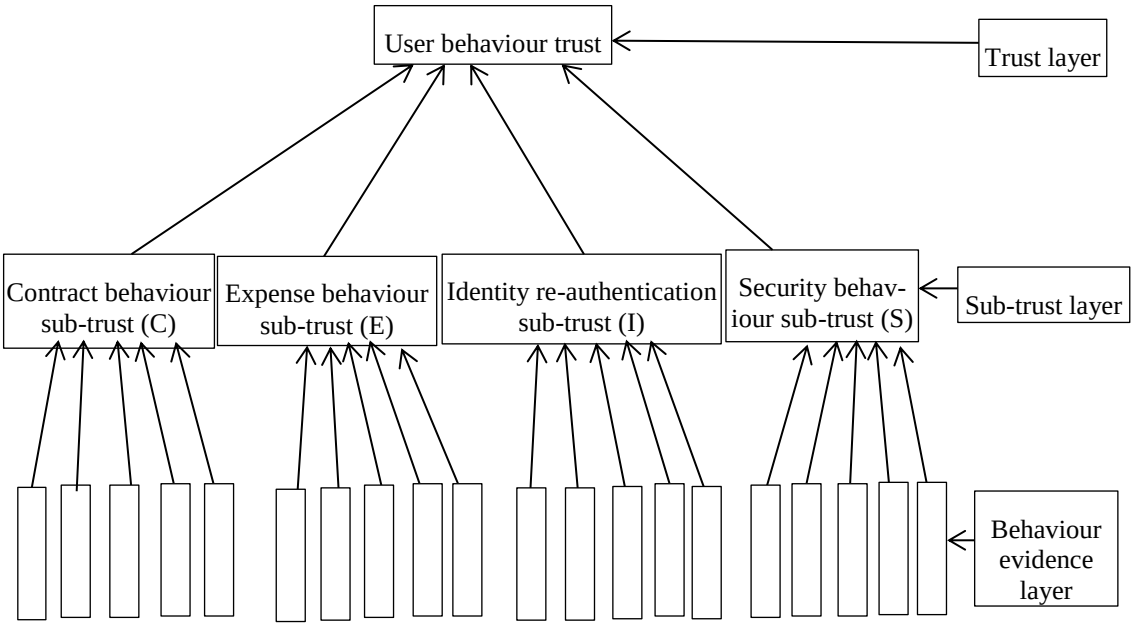


Figure 4.2: user behaviour trust evaluation Hierarchy in cloud computing

Several studies have focused on applying user behaviour profiling of log events of cloud infrastructure services. Krishnan and Chatterjee (2012) proposed an adaptive distributed IDS (ADIDS) based on a novel combination between behaviour and knowledge-based approaches in IDS of the cloud environment. To improve the dynamic system detection for any abnormal usage, the behavioural profiling technique was used to generate users' template from monitoring users' communication requests between VMs, such as time, number of users connected, bandwidth size, port number, and IP address. Then by observing the deviation in the behaviour, a decision can be made; whereas the knowledgebase is to support the detection approach by comparing against the signature of attack patterns. The integration of these two approaches can reduce the false positive rate. Additionally, another aim of the proposed system is designed to send an alert to each cluster node when false alarms are detected from any nodes. This alert is to foster

cooperation between all nodes and to update the database on new attack patterns, as well as to prevent intruders. This collective cooperation serves to protect the infrastructure of cloud computing against any new type of attack and continue to deliver services to users. However, the study was a conceptual work with no practical experiments being conducted.

Similarly, within IDS in the cloud, Doelitzscher et al. (2012) introduced an approach to managing the challenges and limitations in traditional monitoring and intrusion detection techniques when they are applied to the cloud. They focused on auditing the security of cloud infrastructure services through monitoring the suspected change in the resources. Autonomic multi-agents and behaviour analysis techniques were used for auditing an incident detection system, called security audit as a service (SAaaS). This security service was placed inside VMs to collect and analyse the information of these VMs such as frequent infrastructure changes. The aim of their study was to increase cloud security continuously and transparently by informing users about the security incidents of data access and increasing users' trust in cloud services. A normal behaviour for VMs was generated such as the time of start, stop, and delete events of VMs. When these behaviours deviate from the historical behaviours, an alert can be generated and forwarded to the user. Additionally, tenants can access all information details to check status events of their services which were gathered by a security dashboard.

However, this work has several problems. First, although SAaaS is based on multi-agent sensors for collecting different events, the security policies of these

sensors are received from Security Service Level Agreement (SSLA). Therefore, the security policy of the sensor still depends on rules pre-defined by SSLA, which means the detection range will be limited to known attacks. More importantly, a simulated dataset was used to examine the feasibility of this study which might be far from a practical sense. Moreover, using a large number of agents (e.g., initiating agent, killing agent, and moving agent) might increase the communication traffic between the agents and the computational overhead (Khalil et al., 2014). In addition, the prototype is not yet validated for scalable environments because if tenants need to add new services or ask to be protected from a specific type of attack, a problem will occur with autonomous agents. The reason for this problem is the configuration and development of these agents, which is achieved at the beginning of the VM lifecycle. Therefore, there is a need to add global solutions for IDSs in cloud services.

Additionally, Annappaian and Agrawal (2014) proposed an intrusion detection and prevention system (IDPS) based on building user profiles from the regular user usage of cloud services, such as usage timings, durations, access privileges, usage logs, and types of services, where any unusual usage from the generated profile can be detected by cloud providers. The idea of their study is if the comparison value of the historical and current usage behaviours exceeds the threshold value of the system, the intrusion prevention system (IPS) will consider this user's activities as a misuse attack and will raise some questions for further authentication. If this stage is also failed, the activities will be considered as an intruder attack. Finally, this attack will be forwarded to a honey pot and an alert will

be triggered by IDS systems of the cloud provider to stop their service. The authors proposed different states and ranges of the intrusion detection meter, which will help the system to be more accurate in detecting processes, as illustrated in Table 4.8.

Table 4.8: Different Stages and Ranges of Intrusion Detection Meter (Annappaian and Agrawal, 2014)

Stage	Range consider	State consider	Description
1	0-25	Legitimate Usage State	Normal usage
2	25-50	Suspected State	Detect abnormal usage and check against usage profiles
3	50-75	Questionnaire Stage	Usage check against predefined questions
4	75-100	Intruder/Prevention Stage	Honey post to collect some details/Prevent the usage via active response/Inform Vender/Change attacked resource

In the last state (stage 4), a honeypot/prevention state will collect signatures and usage details on this attacker; this hacking information will be sent to the provider to change the policies of the system, update the database of hacking (signatures) for safety actions in the future, and, finally, stop the service to this user. The authors assumed and suggested this proposal might be the good solution for many attacks of IaaS, such as Insider and Denial-of-service (Dos) attacks (when a cyber-attack sends a lot of requests to the system in an attempt to consume enough server recourse to make the system unavailable to legal traffic) and port scanning attacks. However, this study is only theoretical with no practical experiment or evaluation.

In the same year, another proposal for detecting illegal access to the cloud infrastructure was based on a collaborative filtering algorithm. The normal behaviour of cloud user activities (log events) is generated for each user, and then this behaviour profile is compared with other cloud users by applying a cosine similarity

method to collect the score similarity value. The user who accesses the system will be compared with this score value and by using the collaborative filtering algorithm, the ratio of deviations from normal behaviour of cloud users can be determined; if deviation values exceed the system threshold, it can be considered that the user is a masquerade attacker; otherwise, the user will be treated as a legitimate user (Alguliev and Abdullaeva, 2014). However, this work is also solely a theoretical proposal as no experimental work was performed because of the lack of existing datasets.

Apart from authentication/security purposes, another aspect was applied to user behaviour profiles for efficient and resilient cloud management—they are used to manage and monitor the rate and volume of the resource provisioning requirements across cloud services in any server or specific area. This technique allows online monitoring of any change in the cloud resources dynamically through a website, which will consider the predictability of user requirements of cloud resources and improve cloud services. Peoples et al. (2014) did their work based on monitoring the user access and activities with a website (Wisekar), which was hosted in the Indian Institute of Technology Delhi. The experimental results collected the ratio average of visiting the website by visitors within a six-month period based on mobile technology use, country, count, type (new/returning), duration, and type of browser. The evaluation of these results led to better understanding of the behaviour of visitors who use the Wisekar website. The authors mentioned this mechanism can support and determine the high cloud resource demands in the particular places. It can be used for monitoring a user and activity collection to support a security aspect.

Additionally, a number of studies focused on IDS to capture inter-virtual machine traffic and detect a misuse in hypervisor layer (Pandeeswari and Kumar, 2016; Ye et al., 2016; Kashyap et al., 2017). One of the study proposed by the Pandeeswari and Kumar (2016) proposed an anomaly based IDS named as Hypervisor Detector which works on the hypervisor layer to detect any abnormal behaviour in the Hypervisor technique. Hypervisor based IDS helps to monitor and analyse the data communications among variety of virtual machines such as the communications between VMs, VM and hypervisor and between the virtual networks within the hypervisor. Hypervisor Detector uses two hybrid algorithms named as Fuzzy C-Mean clustering and Artificial Neural Network algorithms to increase the accuracy detection rate of the system. The C-Means algorithm was implemented to generate cluster subsets whereas the Artificial Neural Network algorithm was utilised for training each cluster by aggregating modules. They applied the KDD cup dataset, dated back to 1999 for evaluating their works. The experimental results achieved an average detection rate of 97.55% with 3.77% of false alarm rate for detecting intrusions. The dataset was very old compared to the developing techniques of cloud computing nowadays, making it does not match the current cloud infrastructure (Agrawal et al., 2017).

Additionally, Meng et al. (2017) proposed a system called DriftInsight to detect abnormal behaviours in real world for Platform as a Service (PaaS). The idea of this system is to monitor the behaviour of multi components as cloud computing can be a distribution of multiple components that was interacting with each other to achieve a single task or multi tasks. The system involved to deploy more than 100 type of agents which are running in different nodes to monitoring millions of

metrics from the cloud computing. These metrics compute the general performance of CPU, memory, disk and network for each node. By monitoring these nodes continuously, any abnormal behaviour can be detected. The system applied unsupervised clustering algorithm called DBSCAN to convert multi-dimensional metrics which was computed from multi components into single metrics in order to detect abnormal behaviours. The suggested system was evaluated in a commercial large-scale PaaS on real dataset as the authors claimed. However, there are no any details about this data and the authors showed only few statistical graphs without any detection rate which depict the performance of the system.

Recently, Tiwari et al. (2018) tried to develop a system to detect anomaly in cloud Infrastructure as a Service (IaaS) based on user behaviour profiling with the goal of detecting different types of breaches in users VM's resources. The approach was involved to continue monitoring the user usage of the cloud resources including CPU power consumption, network usage data and memory usage. In this study, user's short and long-term behaviours were captured and mark as an anomaly if it deviates from the normal behaviour of user. Methods for e.g. statistical methods such as percentile-based thresholds and moving average, unsupervised classification method through One-Class SVM; and regression via Long Short-Term Memory (LSTM) networks were applied in this study. The unsupervised method achieved an accuracy of 83% to 95%; whereas the regression method achieved 84% of accuracy. However, the statistical methods do not seem to work correctly with complex patterns with accuracy as low as 40%. However, a simulated dataset was used to evaluate the suggested approach which does

not reflect the complete image of the real-world scenarios of users' interactions with cloud infrastructure services.

A study by(Jain and Pandey (2018) proposed a model that can verify the cloud users, based on their daily geographical location usage. When an attacker tries to access the user account without permission, he/she should have all the credentials and the same geographic region where the device is registered previously. Therefore, cloud providers can detect the legitimacy of the current user by comparing the current and historical location of usage. In case if matching result has a large deviation from normal, the cloud provider can restrict the user from using the service. However, the suggested approach was not applied on any type of data and no accuracy rate was achieved.

However, most of the previous works were solely a theoretical proposal as no experimental work was performed because of the lack of existing datasets. Moreover, these studies focused on anomaly detection based IDS from system interactions perspective rather than the users' interactions.

4.6 Discussion

As mentioned previously, the literature review focused on 46 studies using the behavioural profiling technique to protect systems from different types of attacks. However, some of these studies were solely theoretical with no practical experiment; hence, their reliability in practice is not known. Therefore, only studies with practical evaluations on the behaviour profiling technique are presented in Table 4.9.

Table 4.9: Practical Studies of literature Review

Author(s)	Activity	Cli.	#Partici-pants	Performance (%)	Method	Purpose
Moreau et al. (1997)	Telephony	S	#600	DR=90, FRR=10	Supervised Neural Net-works	FD
Burge and Toylor (1997)	Telephony	S	#110	DR=75, FRR=40	Unsupervised Neural Net-work	FD
Samfat and Molva (1997)	Mobility	S	#400	DR=82.5, FRR=40	Distance	FD
	Telephony			DR=80, FRR=30	Rule-base	
Hall et al. (2005)	Mobility	S	#50	DR=50, FRR=50	Instance based learning	IDS
Hilas et al. (2014)	Telephony	S	#5000	Highest DR=80	Genetic Programing method	FD
Ogwueleka (2009)	Telephony	S	#180	F Accept R=3	self-Organizing Map and Probabilistic	FD
Qayyum et al. (2010)	Telephony	S	#300	DR=70	Neural Network	FD
Yazji et al. (2011)	Mobility	S	#100	DR=81	cumulative probability and Marko properties of trajec-tories	IDS
Yazji et al. (2014)	Mobility	S	1-#100 2-#178	DR=94	cumulative probability and Marko properties of trajec-tories	IDS
Subudhi and Panigrahi (2015)	Telephony	S	#94	DR=95,TPR =78, FPR=8	SVM	FD
Shi et al. (2011)	Telephony, SMS, Browsing, Mobility	C	#50	DR=95	Probability	Au
Damopoulos et al. (2011)	Telephony, SMS, Browsing	C	#35	DR=98.5, TPR=99.3 FRR=0.7	Bayesian network , RBF, KNN, Random Forest	Au
Li et al. (2014)	Application Activities	C	#76	EER=9.8	Rule base	Au
Fridman et al. (2015)	Text, App, Web and location	C	#200	EER=3	SVM	Au
Aupy and Clarke (2005)	Way of interaction with PC	C	#21	EER= 7	Neural Network (FF-MLP)	Au
Yazji et al. (2009)	File access activity and net-work events	C	#8	DR=90, FAR=13.7 FRR=11%	K-Means Clustering	Au
Salem and Stolfo (2011)	File access activity	C	#18	FPR=1.1	SVM	Au
Abramson and Aha (2013)	Web Browsing	S	#10	FAR/FRR= 24	SVM	Au

* S=Server, C=Client, DR=Detection Rate, EER=Equal Error Rate, FAR=False Accept Rate, FRR=False Reject Rate, TPR=True Positive Rate, FPR=False Positive Rate, FD=Fraud Detection, Au=Authentication, I=Identification

The table illustrates that a significant number of studies were conducted on investigating the role behavioural profiling plays in enhancing the security within numerous technologies, such as mobile phones, computer systems, networks, and websites. While the earliest studies focused mainly on IDS and fraud detection for telephony systems, the more recent studies have concentrated on the authentication aspect to mitigate misuse. Various activities were collected to build efficient behavioural profiles (e.g., telephony, mobility, application usage, file access activity, network event, and web browsing activity). These activities were collected for both the client and server sides. However, collecting data from the client side might be less reliable because all users' activities are recorded by the device. Therefore, the attacker can learn the user's behaviour pattern or delete, modify, or stop the recorded information if the device is not securely protected.

Most studies used statistical approaches to create features about the work performed by users while interacting with their services or applications. The main advantage of statistical approaches is their ability to simplify or reduce the data particularly on large datasets. For example, Yang (2010) used statistics to generate a feature factor from the raw data of 50,000 users over one year. This can help reduce the effort on a classifier and increase the speed of execution.

Within a similar context of feature collection, a number of studies such as Moreau et al. (1997), Buschkes et al. (1998), Sun et al. (2006), Hall et al. (2005), Hilas and Sahalos (2005), Yazji et al. (2014), and Subudhi and Panigrahi's (2015) depended on a single modality to build user profiles, such as telephony and mobility studies. The highest accuracy achieved by these studies was a 97% true positive

rate (TPR) and 6% false positive rate (FPR). More recent studies have tried to improve this accuracy based on combining multimodal (e.g., telephony, SMS, browsing, and mobility) to build user behaviour profiles. The highest accuracy result achieved was 99.3% of TPR with 0.7% of FPR. It is evident that using multimodal to build user profiles will increase the accuracy of system detection.

However, there are other aspects that can affect a system's accuracy; namely, the volume of data required to be collected for the training and testing stages, which relies on the time of collection and user interaction with services. A study by Chu et al. (2012) showed that the more volume of data collected for training and testing a system, a higher accuracy can be achieved. This can help a classifier build good pattern recognition among users. Additionally, the intervals of data capturing during the test phase can also affect the verification accuracy of the system because of the number of user interactions within the determined time. For example, the system proposed by Salem and Stolfo (2011) can make a decision every two minutes with 1.1 of FPR. Yazji et al. (2009) and Yazji et al.'s (2014) proposed system enables distinguishes between normal use and attacks with an accuracy of approximately 90% every five minutes, whereas within 15 minutes the detection accuracy became 94%. Fridman et al. (2015) have an overall EER of 5% within one minute and 1% after half an hour of the user interactions with his/her smartphone. However, others depend on daily or weekly usage with users who do not have enough frequent activities (Hilas and Sahalos, 2007, Buschkes et al., 1998). Therefore, the resolution of required data concerning time during test stage is a vital aspect in the accuracy of authentication/verification system because if the windows of time is minimised, this might affect data required for

achieving an accurate decision; further, if the time becomes too long, the system might be abused by attackers. These aspects depend on the volume of users' interactions during interacting with service applications; some applications need more time than others to collect enough data for a successful verification process. For example, with keystroke verification, the time of detection can be minimised (might be less than 1 minute) and the data required can be enough to achieve the task whereas other applications might need more time to collect enough data for achieving the system's decision. Therefore, some systems can capture enough data for identifying users within a short time whereas others might need hours or days because of the lack of user interactions. For instance, it is more viable to verify a user has 100 activities a day on s smartphone than that who has merely five activities a day. Also, if the device is not often used, it may store less important information.

Most modern studies have been based on authentication technique to detect abnormal usage to avoid misuse. However, the main problem with the behavioural profiling techniques is stability; users might change their behaviours over time, which might affect the system's accuracy. Therefore, some studies have considered this aspect by renewing templates continuously. For example, Li et al. (2014) proved in their practical experiment that a dynamic profiling has a slightly better accuracy than the static profiling because the dynamic user profiles contain the most recent changing in users' activities. However, this is not an easy task because the renewing template might include illegitimate usage, where an impostor might be accepted by the system over time as the genuine user. Moreover, within the same context of increasing the system's accuracy, adaptive thresholds can

make the system's detection more accurately with systems that their decisions depend on the threshold. Sun et al. (2006) focused on the corresponding detection threshold, which needs to also be changed based on changes in the templates. However, the system should consider that changes in the system threshold might cause system violations by other users or impostors. This is because some users' behaviours might be slightly similar, so when the threshold is changed adaptively (particularly when the system uses a single threshold for all users), the system might accept the samples of other users during the template update. This might lead to accepting impostors as legitimate users by the system over time. Therefore, multi-dynamic thresholds for each user might be needed to make the system more robust for avoiding including impostors' samples with genuine samples during the template renewal phase, particularly for continuous and transparent verification systems.

From a pattern classification perspective, many classification algorithms have been used and different accuracy results were achieved. The main used classifiers were machine learning algorithms (e.g., neural network, Bayesian network, K-means clustering, K-nearest neighbours, Random Forest, Markov chain, genetic algorithm, SVM), as well as rule-based. Damopoulos et al. (2012) showed that the best accuracy algorithm of their work was K-nearest neighbours—99.8% and 99.5% for a true positive rate and accuracy, respectively. The second performance result was given by Random Forest with a true positive rate of 99.8% and 98.9% for accuracy. Moreover, K-means clustering was compared with neural networks, which achieved 98% of DR within 3 months of training data, while neu-

ral networks did not exceed 84% DR (Yazji et al., 2009). The authors also concluded that K-means clustering is often faster and more efficient with large datasets. However, this does not mean K-means clustering is better than neural networks because the accuracy relies on other different factors, such as volume of data and the uniqueness of the selected features. Moreover, the most deployed algorithms were the neural network and SVM, which also achieved a highly accurate result, followed by the rule based methods, and then condition-based algorithms. Although the performance of the most machine learning algorithms is based on statistical or probability techniques (i.e., not a Yes or No answer), they might be more suitable with the unstable behavioural profiling technique. This is attributed to the fact that it is unlikely to extract features from user behaviour that are permanent, where user behaviours are considerably changeable over time. Therefore, it might be more difficult to acquire a better result by using other techniques that are not based on a statistical or probability approach, particularly with an authentication/verification technique applied to re-authenticate users after getting access to a service.

Within the similar context of classification methods, some of these studies attempted to use different classifiers with the same activity although the result was similar. For example, in 2004 and 2006, Sun et al. used mobility activity and applied two classifiers. In addition, in 2005, 2007, and 2014, Hails et al. employed one activity (i.e., telephony) and implemented three different classifiers. However, the accuracy of the results of all these studies was comparable. This means, using different classifiers with the same activity might not have an impact on the system's performance. Other studies attempted to use different activities instead

of a single activity with a single classifier (e.g., neural network). The performance was diverse and better, as illustrated in Table 4.10.

Table 4.10: Different Activities and Results with Neural Network

Study	Behaviour	EER (%)
Aupy and Clarke (2005)	Device usage	7.1
Li et al. (2011)	App. Usage	13.5
	Text message	5.4
	Calls	2.2

As a result, although there is a thought that each problem has to have an appropriate classifier to acquire a better result, the proper selection for the type of features/activities might be more effective to enhance the performance result than merely changing the classifier.

Additionally, with a longer training period, performance also improved (Buschkes et al., 1998; Hilas and Sahalos, 2007; Yazji et al., 2009; and Li et al., 2011). For example, Buschkes et al. (1998) achieved 80% accuracy within five days of the collected dataset, whereas the accuracy increased to 83% with a dataset of 15 days. However, storage and process consumption should also be considered with larger training data samples. Moreover, Li et al. (2014) applied a smoothing function, like a majority voting method, to treat a number of successive applications as a single event. This also led to improving the system's performance, but it took a longer time for the system to make a decision, which made it vulnerable to abuse by an intruder.

4.7 Conclusion

From the preceding sections that have applied the behavioural profiling technique for security aspects, such as IDS, fraud detection, misuse, and authentication, a number of classifiers have been used and many satisfactory accuracy results were achieved, thereby applying different types of activities. These studies have shown and proven the feasibility and usability of using user behaviour profiles to protect a user/system from different types of attacks with respectable performance. Therefore, it is evident that behaviour profiling can offer enriched features and attributes which can be investigated with other technologies, such as cloud computing. To the author's best knowledge, only one practical study was available on using user behaviour profiles for protecting cloud computing services. The study was mainly focused on log events of infrastructure as a service (IaaS). However, it suffered from several limitations; one of these limitations is that no real dataset was used to test their proposed system and the work focused on monitoring the VMs from outside, which means any actions or misuse occurred inside the VMs will not be considered. Therefore, more work on behaviour profiling in cloud computing security is required.

To provide a comprehensive and strong prototype to verify a user's identity in cloud computing services continuously and transparently, user behaviour profiles are required to be built to monitor various real users' activities while interacting with cloud services. SaaS and IaaS will be investigated in this research because, as mentioned in Chapter 2, they can be used by ordinary users compared to PaaS, which is usually used by specialised users such as developers. A real dataset can

be collected from the selected services to test and evaluate the proposed prototype to ensure the capability of adopting this work successfully with considerations taken to a high level of performance, scalability, and interoperability. In this research, there is also a need to consider: (1) selection of suitable features during analysis process, which can give a better result, (2) an appropriate classifier based on the nature of the problem to be tackled, (3) windows of time for the training and testing stages, (4) volume of data that is required for training and/or testing, (5) adaptive and frequent templates update, and (6) dynamic thresholds or multi thresholds for each user. The next two chapters will investigate and explore the feasibility of creating user behaviour profiles from cloud computing services (SaaS and IaaS).

5 User Behavioural in Profiling Software as a Service

5.1 Introduction

The previous chapter discussed using user behaviour profiling with different technologies such as mobile phones, web applications, and computer systems. However, to the best of the author's knowledge, no prior work that uses behavioural profiling has been studied regarding cloud storage services. This chapter presents a feasibility study of using behavioural profiling to verify a user during interaction with the cloud storage service. The main aim of this study is to show the ability of extracting unique features for users through continuously observing users' interaction with these services. These features can be used to generate user behaviour profiles, which will be subsequently applied to detect illegitimate usage (misuse) of services or provide continuous verification to ensure the legitimacy of the current user.

Cloud storage services become particularly attractive for users (both individuals and enterprises) by offering data storage to meet different levels of demand. Customers can upload, download, update, remove, recover, and share data with directly accessing information through online web applications from anywhere at any time. The flexibility, accessibility, simplicity, efficiency, scalability, and pay-as-you-go feature offered by cloud providers made the number of subscribers increase rapidly (Forbes, 2015). There are widely popular cloud computing services, which can be chosen for investigation in this study. Dropbox, Google Drive, One

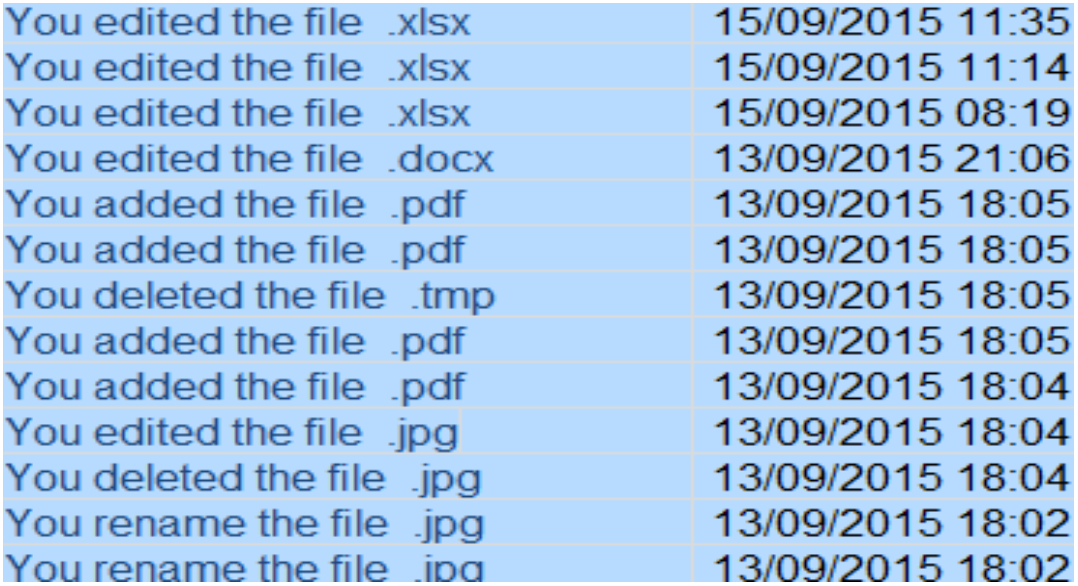
Drive, and Box are all examples of widely popular cloud storage services. Dropbox was chosen for this research because it is one of the most popular cloud storage services (Erin Griffith 2014, CloudRAIL 2017) and, importantly, it provides simple access to users' interactions records. A series of experiments on a private collected dataset have been conducted to understand the degree to which behavioural profiling can be applied to increase the security of cloud storage services.

5.2 Methodology

5.2.1 Data Collection

Data collection for user activity in cloud computing services is challenging because of the unwillingness of cloud providers to share it; one of their reasons is to protect customers' privacy. Consequently, the way by which the real data of users' activities in cloud computing services can be collected is not straightforward. In addition, to the best of the author's knowledge and investigation, no public dataset on user's cloud activities is available. As such, Dropbox was selected for the collection of a private real dataset because of a number of reasons. Firstly, Dropbox is a well-known cloud storage service with more than 500 million users, 300,000 businesses on Dropbox Business, and 400 billion total pieces of content uploaded daily (Dropbox 2018). More importantly, Dropbox provides users' interactions within the Dropbox web application for six months and those logs can easily be obtained from the Dropbox web interface rather than waiting six months to collect them. The author collected private real dataset from 30 participants. This number was selected based on a practical reality behind the required participants to provide data. The analysis of previous studies on biometrics has also

shown that 30 is the minimum group size used to evaluate the feasibility of studies. However, care still has to be taken in generalising a result. In terms of the quantity of interactions over a specific period, it would be inductive. Therefore, a PhD research group was selected to collect this dataset for many reasons; firstly, they were selected based on their availability and the fact they used Drobox for storing their work. Secondly, they were all doing similar kinds of work within the same working hours. This would make the research problem difficult and challenging because of similar user patterns. Therefore, if the behaviour profiling technique worked successfully with these users, it would work easily with other types of groups. Therefore, a private dataset was collected from a cloud storage service (Dropbox) containing real user interactions of 30 participants over a six-month period from 02/09/2015 to 02/03/2016 (totalling 91,371 log entries). More importantly, the author assumed that this dataset does not contain any malicious activity from either under human being (legitimate users who interact with the service) or deal with malware. The data has been anonymised to protect the participants' privacy and ethical approval was sought and obtained from the authors' institution (Appendix B). Figure 5.1 demonstrates a sample of user interactions with Dropbox.



You edited the file .xlsx	15/09/2015 11:35
You edited the file .xlsx	15/09/2015 11:14
You edited the file .xlsx	15/09/2015 08:19
You edited the file .docx	13/09/2015 21:06
You added the file .pdf	13/09/2015 18:05
You added the file .pdf	13/09/2015 18:05
You deleted the file .tmp	13/09/2015 18:05
You added the file .pdf	13/09/2015 18:05
You added the file .pdf	13/09/2015 18:04
You edited the file .jpg	13/09/2015 18:04
You deleted the file .jpg	13/09/2015 18:04
You rename the file .jpg	13/09/2015 18:02
You rename the file .jpg	13/09/2015 18:02

Figure 5.1: User Activity with Dropbox

To extract features from the available dataset to build a user behaviour profile, Figure 5.1 shows that a number of user's activities can be investigated. These features are:

- Events such as 'Add', 'Delete', 'Edit', 'Move', and 'Rename'.
- File types such as 'docx', 'xlsx', 'pdf' and 'tmp'.
- Timestamps such as date and time of access .

These standard features can be expected to see them with any cloud storage system such as Google Drive and One Drive. These features might help to extract unique features to generate user behaviour profiles. Table 5.1 below shows some of the main user activities extracted from the above figure.

Table 5.1: User Dropbox Activities

Event	File Type	Time and Date Stamp
Edit	xlsx	15/09/2015 11:35
Edit	xlsx	15/09/2015 11:14
Edit	xlsx	15/09/2015 08:19
Edit	docx	13/09/2015 21:06
Add	pdf	13/09/2015 18:05
Add	pdf	13/09/2015 18:05
Delete	tmp	13/09/2015 18:05
Add	tmp	13/09/2015 18:05
Add	pdf	13/09/2015 18:04
Add	pdf	13/09/2015 18:04
Edit	jpg	13/09/2015 18:04
Delete	jpg	13/09/2015 18:04
Rename	jpg	13/09/2015 18:02

From Table 5.1, there are many ways that can be investigated for extracting some features for Dropbox users, which might be unique for each user and can help to build a user behaviour profile. For example, some users access their Dropbox accounts to read, rename, or download files, whereas others might mostly edit or upload files to their accounts. Moreover, these files can have various extensions (e.g., .pdf, .doc, .xls, and .jpg) as users might deal with specific types of files. The time, duration, and date of access might also be another factor that can be used to discriminate users. For example, when an impostor accesses a user's account, they may choose different types of operations which the owner might not use, or the date and time of access to the account might be different, such as deleting files on Sunday after 12 PM when the legitimate user might not use his/her Dropbox account at that time on that day. Consequently, various user behaviours within Dropbox activities can be investigated to verify and discriminate between authorised and unauthorised users through creating a unique pattern for each user that can be observed while interacting with the service.

To generate user behaviour profiles, the number of user interactions or activities should be enough to obtain user's normal usage patterns. For example, it could be difficult to build user behaviour profiles from one or two interactions within a day of usage. Therefore, there is a need to know the number of these interactions/activities for users and the nature of this dataset. Table 5.2 below demonstrates the data aggregated for 30 users during the six months of usage. Some of the file types have not been selected because of low frequency usage, such as them being used once or twice over 6 months, which will not give an informed understanding of users' usage pattern.

Table 5.2: Overview of Dropbox dataset

Number of participants	30
Number of interactions	91,371
Number of unique events	5
Number of unique file types	108
Average of user interactions per day	18.73

Firstly, the interactions in Table 5.2 mean one of these events: "add", "delete", "edit", "move", "rename" upon one of a file type. The table shows the average user interactions based on daily usage can be considered rich enough for meaningful analysis. However, there is a need to see the possibility of discriminating users based on these interactions. Moreover, there is also need to pinpoint user usage individually to explore the nature of their interactions deeply and compare the patterns of these interactions with other users. This can help the investigator find patterns for each user over time and determine dissimilarities among other users. Keeping similar user patterns over time with different usage from others, will help to build sufficient user behaviour profiles.

5.2.2 Experimental Procedure

As previously mentioned, this study is aimed at focusing on understanding the degree to which behaviour profiling can be used to verify individuals within cloud storage services—understanding whether a user is legitimate provides a basis for the system to respond. Therefore, four experiments were conducted on users of Dropbox to examine different factors. These experiments are listed bellow:

- **Experiment 1** use a descriptive statistic method to evaluate and analyse features that are presented by the dataset through extracting unique patterns to discriminate individuals.
- **Experiment 2** to investigate the nature of different classification approaches to explore how performance would be affected by changing the configuration settings. The findings of this experiment would also help identify the optimal classifier to solve the behaviour profile issue within cloud storage services. The experiment used the 66/34 splitting for the training and testing of data with a random selection for the samples.

- **Experiment 3** to explore the impact of the volume of data for training and testing on the system's performance. In addition to the 66/34 splitting, 50/50 and 80/20 splitting approaches are also used for training/testing with random sample selection. Regarding the classifier, the classification algorithm that achieves the best performance from the second experiment was chosen for this experiment. The comparison between the accuracy of the result of each data volume gave a better understanding of the nature of user behaviour profiles.

- **Experiment 4** to understand the effect of time series rather than random sample selection on the accuracy of a decision. To understand the effect of the two metrics on performance, the similar volume of data for the training and testing sets of the third experiment is applied. The accuracy of each volume is compared with the accuracy of the volume of the previous experiment.

Based on the literature survey on behaviour profiling in Chapter 4, four supervised learning classification methods were identified as they achieved good performance with various domains and their datasets were similar to our dataset. As mentioned previously, no single classification method can solve all classification problems. Therefore, these four supervised machine learning algorithms were applied to the given problem to select the optimal classifier for the best performance. These classifiers are: Support Vector Machine (SVM), Random Forest (RF), Feed-Forward Multi-Layer Perceptron (FF MLP) neural network, and Classification and Regression Trees (CART). The first three classifiers were selected based on a high accuracy that was achieved with the previous studies of various other technologies while the fourth method was selected based on Wu et al.'s (2008) study conducted on different classification algorithms. The CART was a one of the best algorithms that achieved the highest performance.

MATLAB was used as the investigation platform for the experiment because it comes with toolboxes (libraries), such as statistical and classification toolboxes. Therefore, a MATLAB script can be written to call these libraries easily to examine any statistical or classification method.

As the following information is available in the given dataset: timestamp of the action (day, hour, and minute), the file type (e.g., .pdf, .jpg, and .docx), and user's action (i.e., add, edit, delete, move, and rename), this information was selected as the main feature set for implementation in the classification experiments, which could provide a good level of pattern recognition amoung users.

To make those features acceptable using classification algorithms as inputs, the symbolic-value attributes (e.g., file type and user action) were enumerated into numeric-value attributes. Then the linear normalization technique was applied, which divides each feature of a vector by the maximum value of that vector to convert the scale of the selected features into the range of 0-1 (Sola and Sevilla, 1997).

The records of each user were divided into two sets: the first set was used to generate a profile for the training stage while the second set was used to evaluate the classifiers' performance at the testing stage. For evaluation, each user's data is considered as legitimate data for the corresponding user, whereas it is treated as imposter data for other users (i.e., as the problem is the user's personal data, referred to as normal behaviour data, and other users' data represent potential abnormal behaviours). Also, equal error rate (EER) is used to evaluate the performance of classification algorithms (Jain et al., 2006).

5.3 Experimental Results

5.3.1 Descriptive Statistics

The preliminary feature analysis implemented a descriptive statistic method to analyse and understand the underlying dataset to determine whether there are unique patterns that can be used to discriminate individuals (Sallehuddin et al., 2015). Selecting an effective or an optimum set of features is a critical and significantly important process because it will subsequently affect pattern classification and the system's performance (Nguyen and Torre 2010). A period of six months was applied to measure the degree of stability of these features over this period. As user behaviour tends to change over time, this long period can help us understand users' patterns more deeply and suggest better techniques to deal with these changes. Intra and inter classes variance was applied to explain users' patterns. The intra classes variances means the similarity of user's pattern/usage should be comparable over time. The value of user intra-classes variance should be small, whereas the inter-classes variance should be large. This means the user's pattern is different from other users' usage. Therefore, the intra and inter classes variance was applied to explore the stability of user's patterns over time and how they are different from each other. These factors can help a classifier distinguish between users and make an accurate decision.

Generally, an initial analysis was made over the dataset to determine the level of usage (low, median, and high interactions) for each user. Some users might interact with Dropbox on one or two days during the week while other users might be active all week. Moreover, the number of interactions during the day of usage

could be different; some users might have high interactions while other users may have less. This could be used as discriminative information among users and from a statistical perspective, it can help divide the users as groups and look on the user data of each group more deeply.

Based on the literature review, some systems can capture enough data for identifying users within a short period such as seconds or minutes, whereas others might need hours or days because of the lack of user interactions. It is arguable that Dropbox might not be used quite often like keystroke as most users use it to store their data occasionally. Therefore, the investigation explored the users' interaction within daily and weekly timeframes to determine any possibility to separate the users as groups, as shown in Figures 5.2 and 5.3.

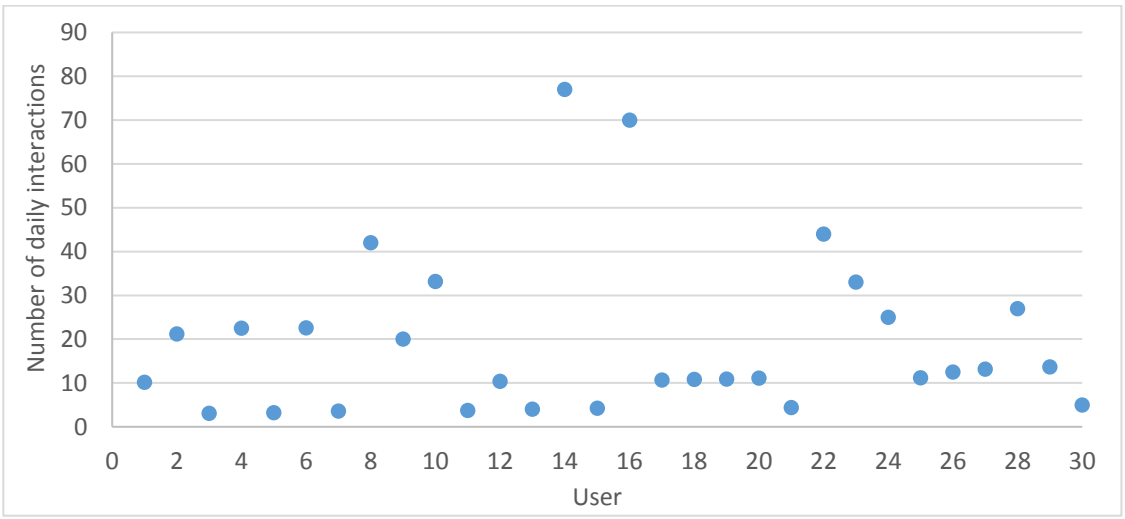


Figure 5.2: Average of users' daily interactions

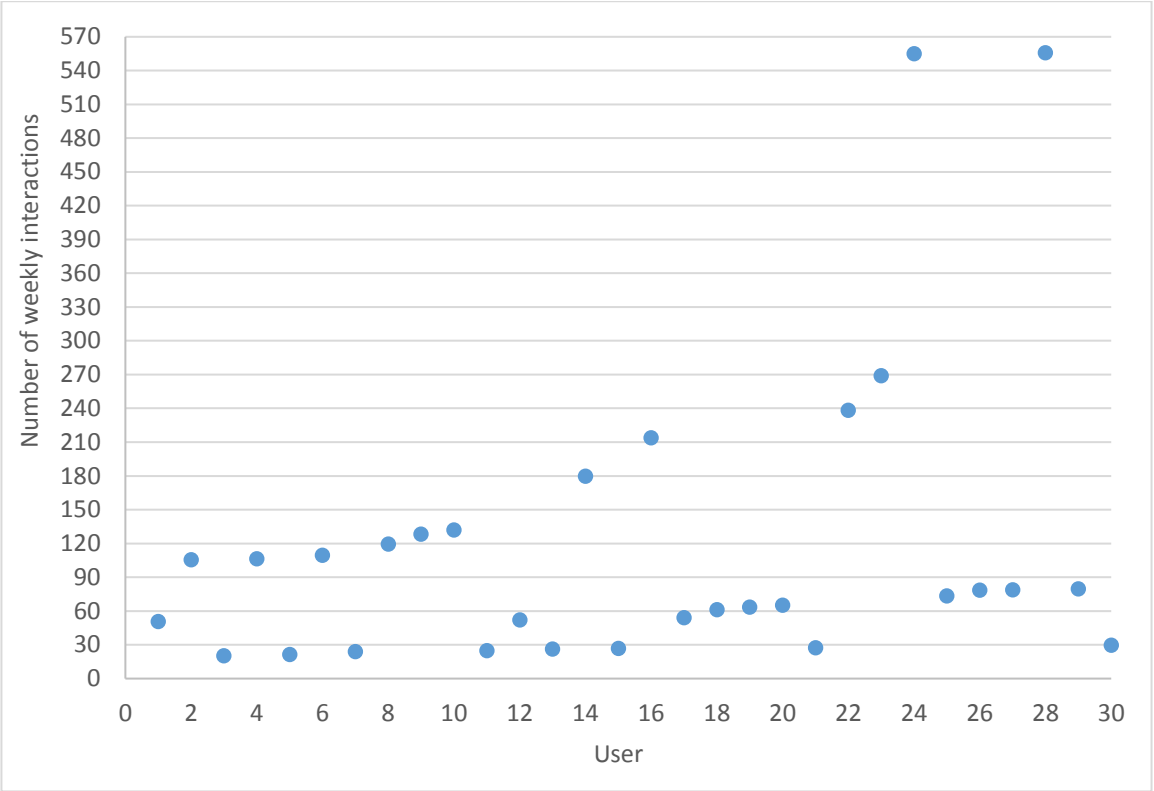


Figure 5.3: Average of weekly users' interactions

From Figure 5.2 and Figure 5.3, it quickly becomes evident that it could be quite challenging to deal with the given population as one group because some users have few interactions and other users have many. Therefore, based on daily and weekly user interactions, users were divided as three groups. The first group included twelve users who had a high number of interactions, where their interactions were equal or greater than 20 interactions based on daily usage and equal or greater than 105 based on weekly usage. The second group (the medium interaction) included 10 users who had from 10 to 20 interactions based on daily interactions and from 50 to 80 based on weekly usage. The last group was the low interaction group and had eight users who had interactions equal or less than

five interactions an average daily and equal or less than 30 interactions an average weekly. Table 5.3 summarises all these groups with their users.

Table 5.3: Users categories based upon their interactions

Level of usage	Participants' number
High usage	2, 4, 6, 8, 9, 10, 14, 16, 22, 23, 24, 28
Medium usage	1, 12, 17, 18, 19, 20, 25, 26, 27, 29
Low usage	3, 5, 7, 11, 13, 15, 21, 30

From a descriptive statistics perspective, it is arguable that the users can be distinguished as groups based on the amount of interactions. However, to further understand the nature of the data, a number of analyses were performed daily and weekly. The first group was analysed based on daily usage, as their users had a high usage, which might provide rich information to distinguish among users. The other two groups were analysed based on weekly usage, as they have low interactions.

5.3.1.1 Group of High Usage

As previously mentioned, 12 users were selected for the group with a high number of interactions. Based on available information in the selected dataset, user events and file types with their date and time stamps were examined to explore whether any discriminative information could be used to differentiate users within this group. Firstly, the volume of this information was explored to determine whether there were enough interactions that could help with this analysis. Table 5.4 below shows basic information about the selected features used by the 12 users during the six-month period.

Table 5.4: Number of users' events and file types

Activity	Frequency
Add	23,491
Edit	40,654
Delete	3,175
Move	1,081
Rename	2,584
Unique File Types	46

Table 5.4 shows the number of interactions of this group could be investigated to explore the discriminatory information for users based on daily usage. In this analysis, two common events were also selected ('Add', 'Edit') as they represent nearly 90% of users' interactions, which could give a good consistency pattern for each user over the chosen period. Mean and standard deviation were implemented to grasp the degree of the intra and inter-classes variance among the users, as illustrated in Figure 5.4.

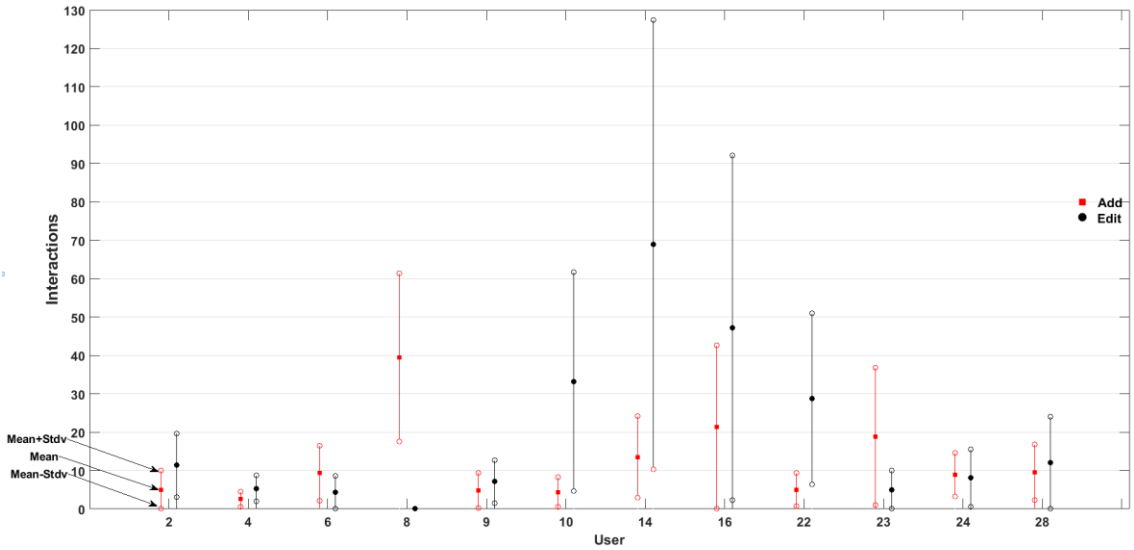


Figure 5.4: Users mean & standard deviation

Figure 5.4 shows that each user has different average usage, as well as some users have a good standard deviation such as users 2, 4, 9,10, 22, and 24 for

‘Add’ event and User 23 for ‘Edit’ event, which means they have a closer pattern of usage over the weeks. The different averages with consistent patterns of usage could help build good user behaviour profiles. However, the users’ standard deviation shows most of the users overlapped with each other regarding the average interactions, which can cause a problem within verification process.

Another feature that could be investigated to see the possibility of identifying users is file types. If no two users used the same file type, this could be used as a unique feature to distinguish between users. Figure 5.5 demonstrates the users’ usage for various file types during the six months.

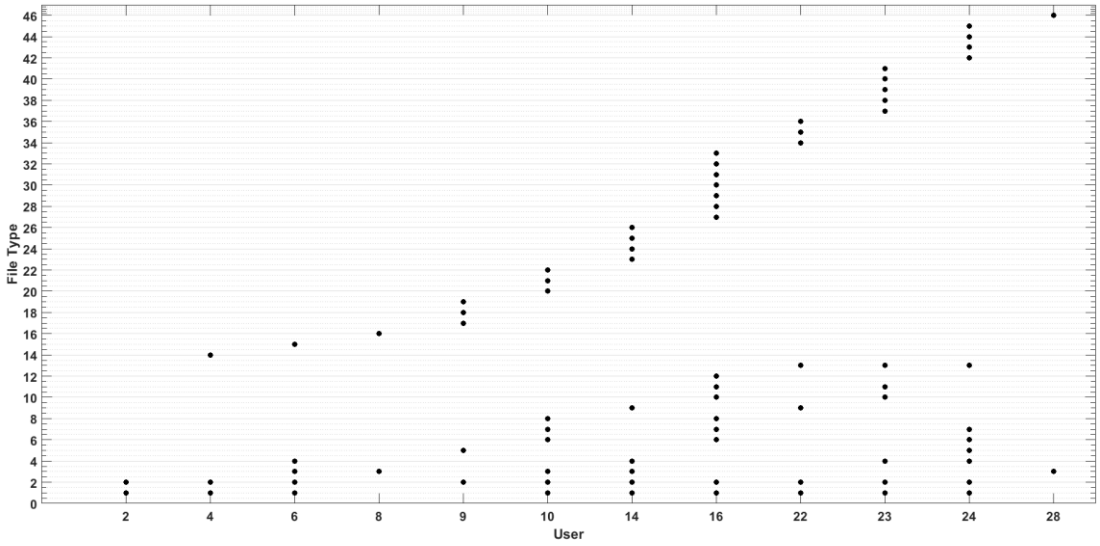


Figure 5.5: Users with their file types’ usage

Figure 5.5 shows that all users have at least one unique file type except User 2. As a result, using this feature can help distinguish easily between the users i.e., inter-class variance is achieved. For example, usage a proportion of unique file types for users 23 and 28 is high: 64% and 67% of total usage, respectively, as

shown in Table 5.5. Therefore, they could be discriminated easily from other users based on the unique file type feature. From an intra-classes variance respective, Table 5.5 also demonstrates the frequent usage of these unique file types across the all given days.

Table 5.5: Usage of Users unique file types over six months

User	Proportion of total usage of unique file types	Frequency of unique file type/Total usage days
4	22%	96/149
6	21%	126/156
8	9%	89/114
9	43%	113/154
10	47%	76/119
14	46%	150/166
16	25%	107/143
22	6%	64/129
23	64%	109/129
24	31%	66/167
28	67%	132/151

Therefore, the intra-classes variance is also achieved in all users that appeared in the table because the frequency of a unique file type existed in more than half of the total usage over the chosen period. However, Table 5.5 also shows that for other users, such as users 8 and 22, most of their file types are similar to the other users. This could affect the verification process to distinguish these users based on this feature. As seen in Figure 5.6, the users shared a number of common file types such as .docx, .pdf, .xlsx, .pptx, and .jpg . For example, User 2 had no unique file type; he/she mainly used two file types (.docx and .pdf), which are shared with most users. Therefore, more investigation is required to determine the possibility of distinguishing the usage of these shared files.

Two types of investigations were applied to three common file types (.pdf, .docx, and .jpg) used by most of the users. Users who used at least one of these common file types were selected in these investigations. The first investigation was by calculating the mean and standard deviation of the usage of these file types, as shown in Figure 5.6.

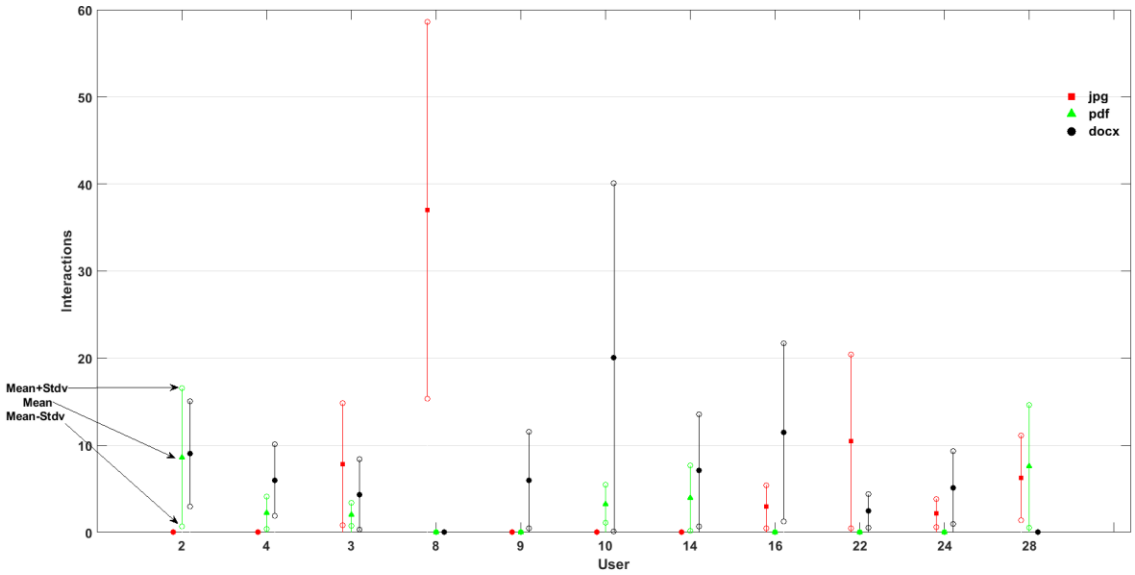


Figure 5.6: Users mean & standard deviation

Figure 5.6 shows that the most of the users had a different average usage for these shared files. Moreover, some users had a good standard deviation such as users 16 and 24 for '.jpg' and users 4, 6, 10 for '.pdf', which means they have a closer pattern of usage during the chosen period. Different average usage could be helped to identify users and consistent usage could help identify users individually. However, many of the users' standard deviations overlapped with each other such as users 4 and 6, who had a closer average of usage for the file types '.pdf' and '.docx'. This could increase the possibility of the classifier's inability to discriminate between them. This would affect the system's performance.

More importantly, from Figure 5.4 and Figure 5.6, it can be concluded that User 8 used Dropbox as a backup to upload only images '.jpg' every time. Although he/she has a high standard derivation, the average of use was higher than all other users without overlapping. Therefore, this user could be easily distinguished from other users based on these features (event and file type): 'Upload' and 'jpg'.

The number of file types used by users during a day over the given period could be investigated as another feature to distinguish among users. Figure 5.7 below shows the mean and standard deviations of the average of daily usage for the number of file types.

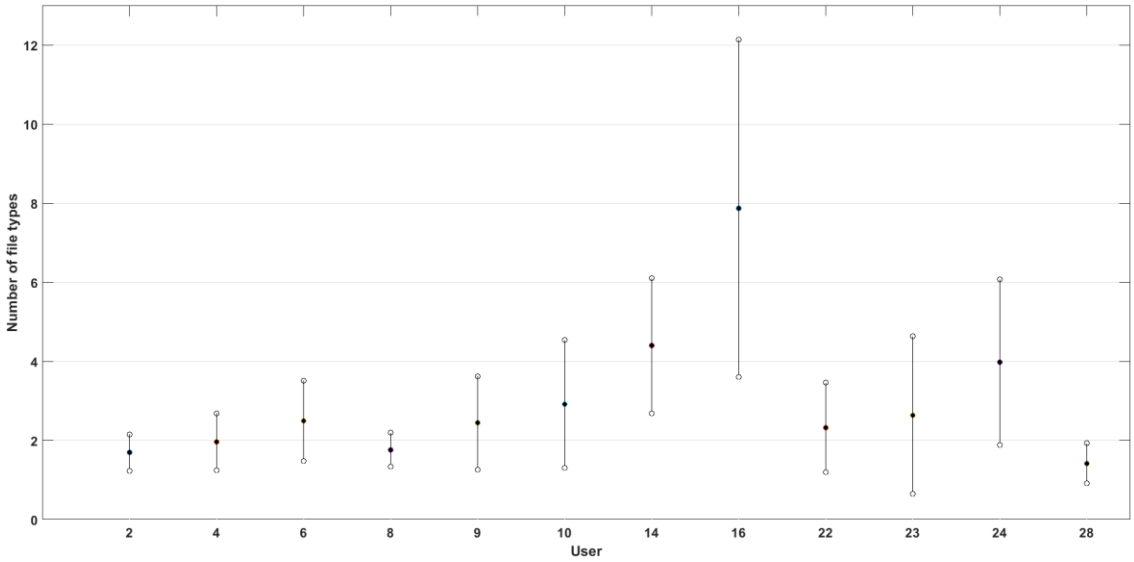


Figure 5.7: Users mean & standard deviation for the average number of the daily file types

From the above figure it can be seen that some users, such as users 14, 16 and 24, had a different average of the number of file types used during a day, which could support the recognition process. In addition, some users have a good standard deviation, such as users 2, 8 and 28. This could be useful for identifying an

individual. However, it can be noticed that there is a degree of overlapping between the users, which might negatively affect the system's performance.

The average of usage during weekdays could be another feature that might help to discriminate the users. Mean and standard deviation of the weekday usage were calculated. Six users were selected who have a high percentage of sharing the same file types to examine this feature, as shown in Figure 5.8.

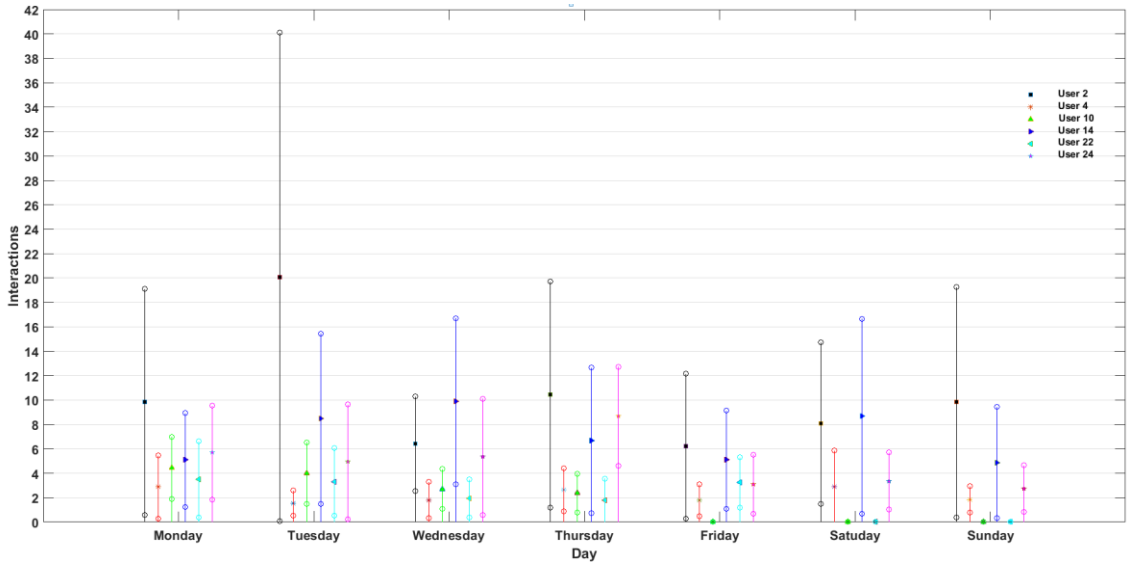


Figure 5.8: Users mean & standard deviation for the average of weekday's usage

Figure 5.8 demonstrates that each user has a different average of usage during weekdays over the six months. Moreover, some users did not use Dropbox on a particular day; for instance, User 10 did not use their account on Friday, Saturday, and Sunday; similarly User 22 also did not use Dropbox on Saturday and Sunday. This could be used to distinguish those users based on daily usage. Further, any access to their accounts on these days could be considered as abnormal activity.

However, most users' standard deviation overlapped, which might affect the distinguishing process.

Time stamps could be investigated as another feature to discriminate between the users of this group. Night, morning, and evening times were divided to examine the average of their use over the given period. Mean and standard deviation were applied to compute the average of use for these three times, as illustrated in Figure 5.9.

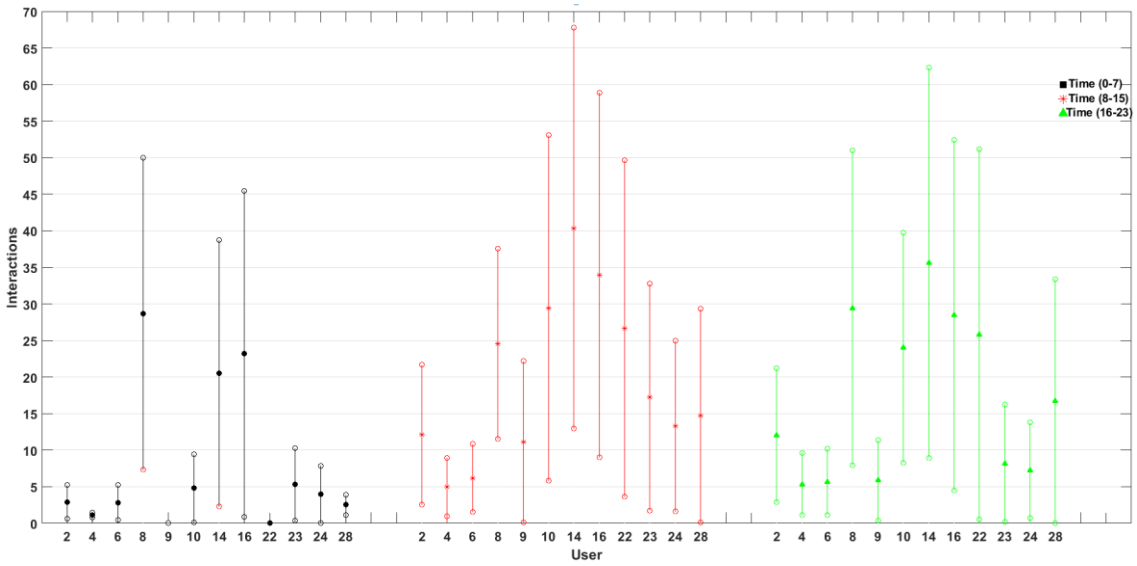


Figure 5.9: Users mean & standard deviation for the three time periods

It can be noted from the above figure there is a different average of use for each user over the six months based on the three times during the day. Some users did not interact with their account during the night, such as users 9 and 22. Further, some users have a good standard deviation across all the selected times such as users 4 and 6. All these aspects could be implemented to increase the discrimination among the users.

5.3.1.2 Group of Medium Usage

As mentioned previously, this group includes 10 users who have medium usage compared with the other two groups. The events and file types with their date and time stamps are the main features in the given dataset. Table 5.6 below shows the basic information about these features. This information contains the number of interactions of each event with 28 unique names of file types used by the 10 users during the six-month period.

Table 5.6: Dropbox events and file types of medium usage group

Activity	Frequency
Add	5,181
Edit	10,186
Delete	1,008
Move	511
Rename	712
Unique File Types	28

Table 5.6 shows the number of interactions could be enough to generate user behaviour profiles for the 10 users. Therefore, several investigations were applied to determine if there was enough discriminative information between these users. The following features were available and analysed: type of events (“add”, “edit”, “delete”, “move” and “rename”), file types (“doc”, “ppt”, “xls”, “jpeg”, “pdf”, “docx”, “xlsx”, “rtf”, “gmf”, “myi”, “c”, “out”, “pqc”, “png”, “pptx”, “enl”, “mov”, “jpg”, “gif”, “aux”, “r”, “tex”, “log”, “gz”, “txt”, “bib”, “blg” and “gms”), time of user’s activity.

Two events (‘Add’, ‘Edit’) were selected to analyse the event types being used as they represent 87% of users’ interactions, as shown in Table 5.6, which could give a good consistency pattern for each user over the chosen period. Mean and

standard deviation was implemented on the use of these two events to find the degree of similarity in usage patterns for each individual and dissimilarity among others, i.e., intra and inter-class variance, as illustrated in Figure 5.10.

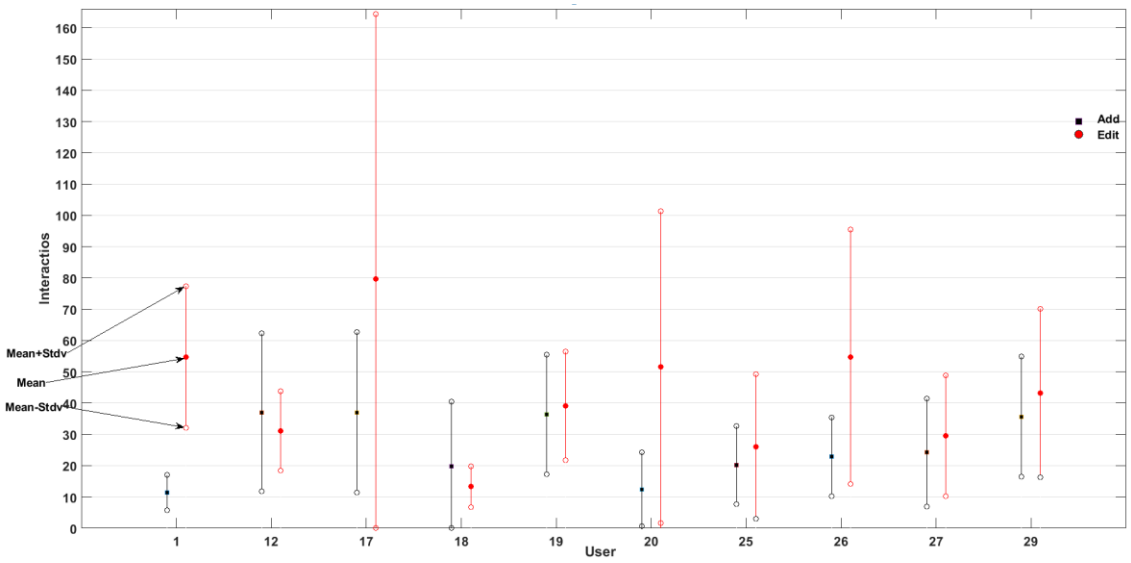


Figure 5.10: User mean & standard deviation

Figure 5.10 shows that some users had a different average usage based on the selected events. For example, User 1's usage for the 'Add' event and User 18's for the 'Edit' event was different from most other users, which could be useful for separating users from others. However, the standard deviations of the most users showed also that users overlapped with each other in usage of these events, particularly users 17 and 20 for the 'Edit' event. This might lead to a poor distinguishing between the users. Therefore, it would be difficult to discriminate users only based on the event types they used over a chosen period.

File types could be another feature to support the possibility of discrimination the users; if no two users shared the same file type, this could be used as a positive feature for classifying Dropbox users. Figure 5.11 below demonstrates that most

users had various usage of unique file types during the six months; at least one file type is unique for each user except User 25. Therefore, inter-class variance was achieved.

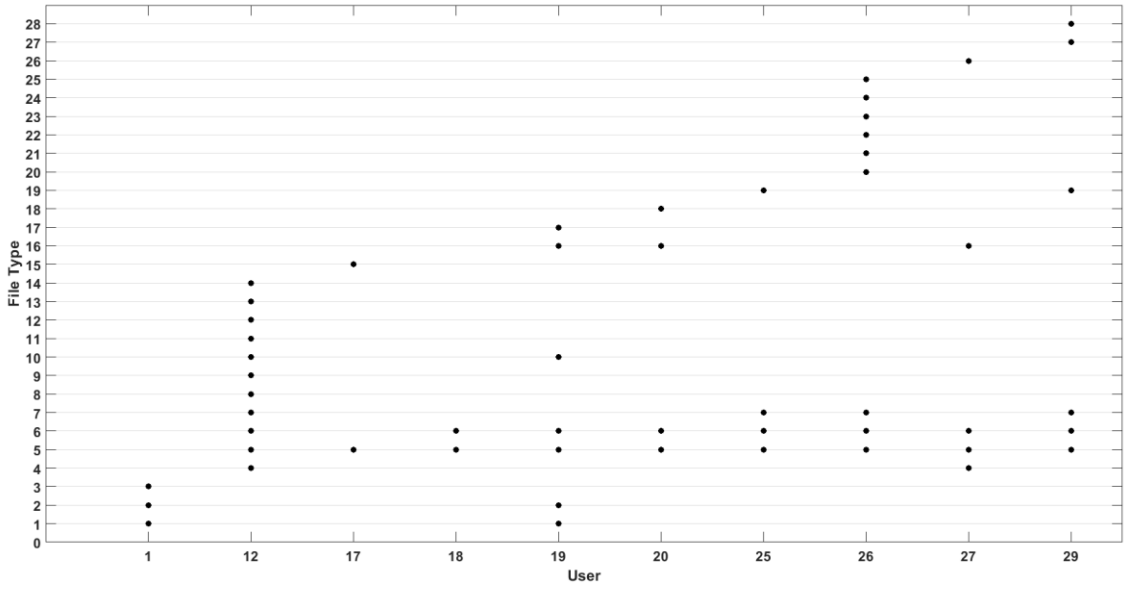


Figure 5.11: Users with their file types

Additionally, Figure 5.12 shows that the intra-class variance was also achieved for the most users because the consistency of usage for the file types of each individual remained comparable during the weekly usage. This can give an indication that users can be distinguished based on the pattern of file type usage.

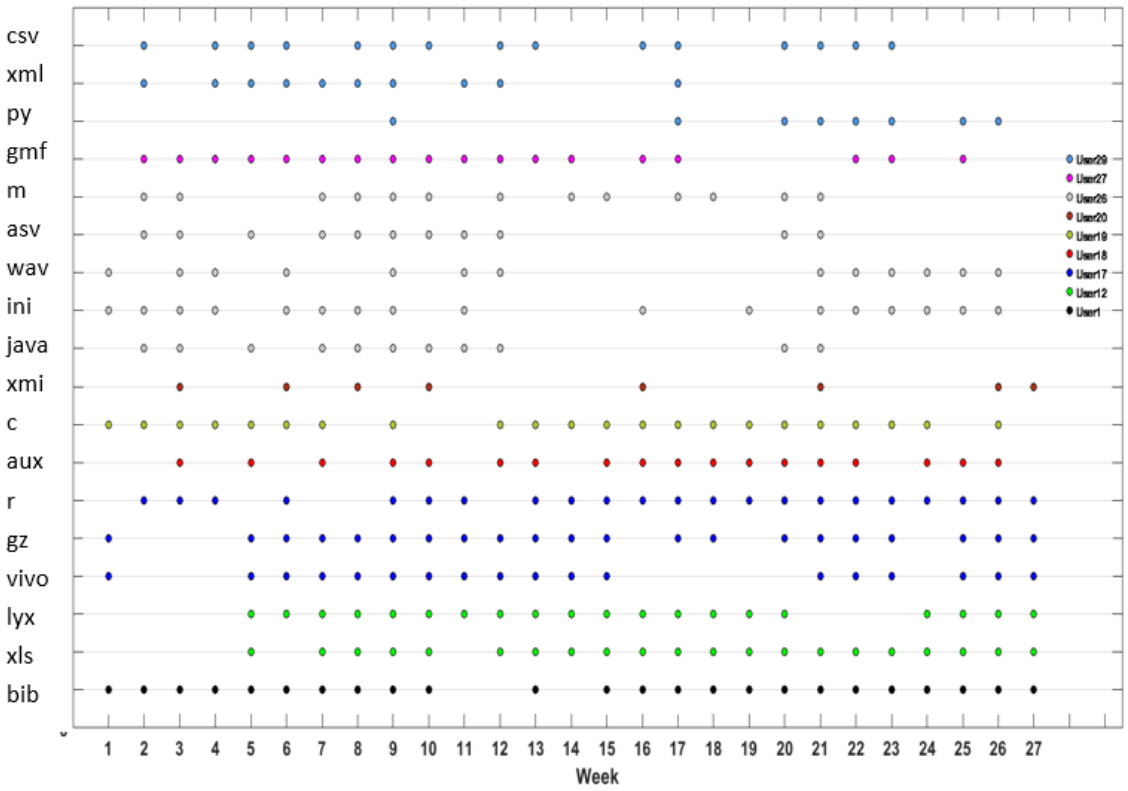


Figure 5.12: Users unique file types' usage across six months

Therefore, the file type may be a strong feature that can help distinguish users if users did not use the same file types. However, Figure 5.12 shows that some users shared a number of common file types such as .docx, .pdf, .xlsx, .pptx. For example, User 20 shared the file type (.docx) among most of the users, which was 85% of their usage. Consequently, it would be difficult to discriminate some users solely based on the file types. Therefore, further investigation is needed to find the possibility of discrimination between users that share the same file types.

Therefore, the two most frequent file types ('docx' and 'pdf') among most users were selected for further investigation to examine another discriminatory factor between the users. The first investigation was by calculating the mean and standard deviation of the usage of these file types, as shown in Figure 5.13.

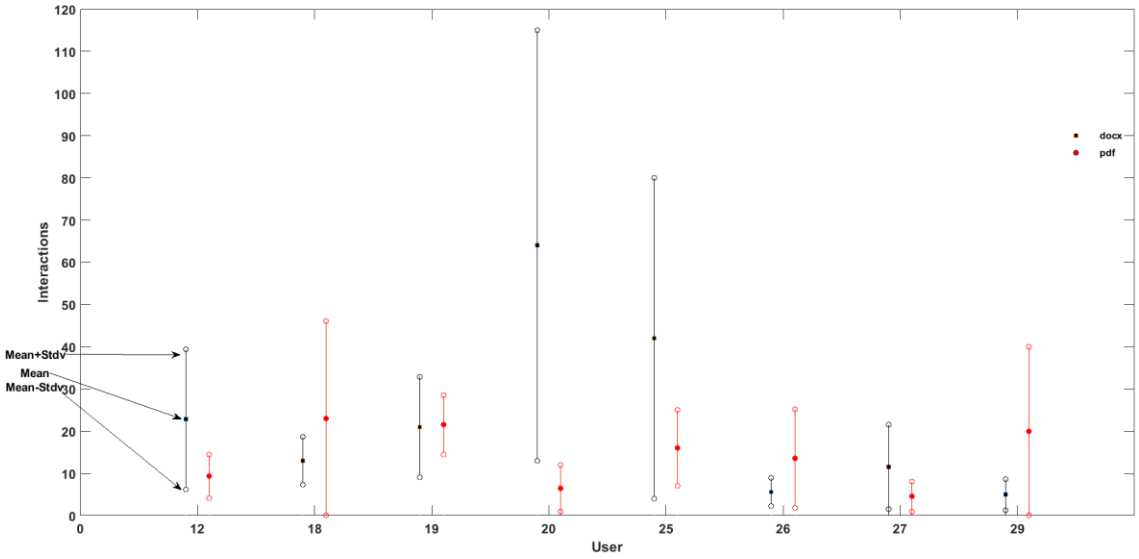


Figure 5.13: Users mean & standard deviation

Figure 5.13 shows that 8 out of 10 users in this group shared the file types (docx and pdf). However, it can be seen that most of the users had a different average of usage for these file types. Moreover, some users had a good standard deviation, such as users 12, 19, 20 and 27 for '.pdf' and users 18, 26, 29 for '.docx', which means they had a close pattern of usage across the given period. This could help with identifying users individually. However, comparing the standard deviations of all users can show that the usage of the most users can overlap with each other. This would affect the ability of a classifier to distinguish between them. However, what has not been considered is a classifier can look at the users' features from multi-dimensional space. Therefore, if examining the use of both file types together, a better-distinguished pattern between the users can be found. For instance, User 12 had higher usage for .docx than User 18 did, whereas User 18 had higher usage for .pdf than User 12 did.

The average usage hour for the above file types ('.docx' and '.pdf') over the chosen period could be also another feature for identifying users. Figure 5.14 shows the examination of this feature by calculating the mean and standard deviation for the usage of these file types.

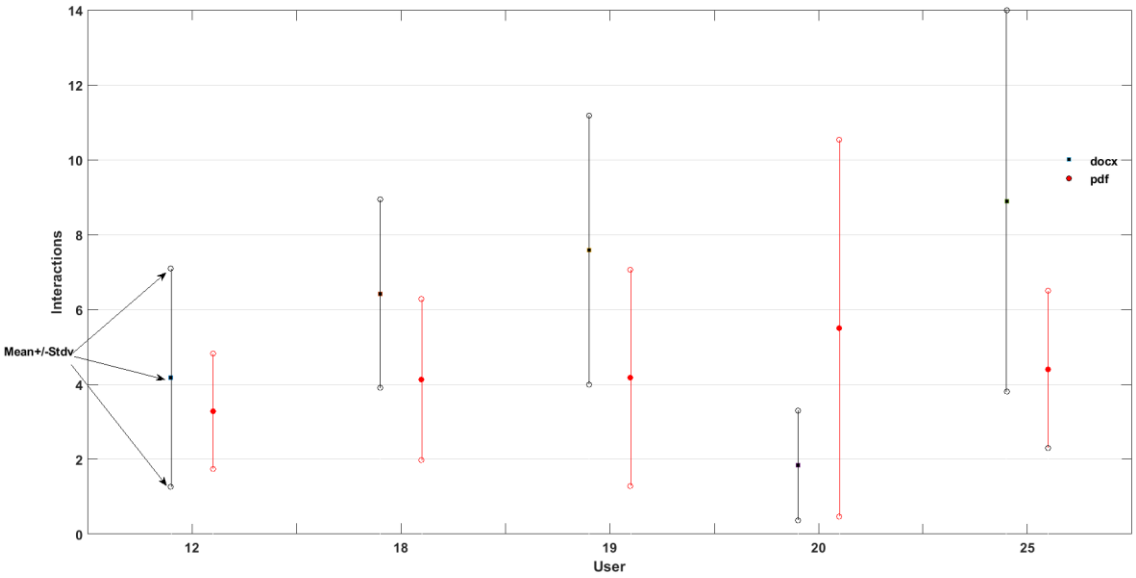


Figure 5.14: Users mean & standard deviation for the average usage of the hourly file types

Figure 5.14 shows that the average usage hour during the weeks for the .docx and .pdf)is different from a user to another person, and some users have a good standard deviation, such as User 12 for the .pdf and User 20 for the .docx. This feature could be applied to separate between users based on hours of usage and the file types being used.

5.3.1.3 Group of Low Usage

Based on the initial analysis of the available dataset, eight users were considered for this group who had limited interactions. The number of interactions for these eight users was 2,788 interactions over the six months. Therefore, the average number of interactions for each user was small, which could be difficult regarding

applying similar previous criteria for finding a good consistent pattern for each user. Consequently, there is a need to find another method that the system can apply it to protect the users of this group.

The idea of anomaly detection profiling could be implemented to protect the system from attacks. The idea is to extract features from all users to build rules for the system and if any user violates these rules, the system can make a decision regarding that user. Therefore, behavioural profiling can be built for the system based on most command features that are shared among users, such as the average interactions within a specific time, average file types that are used, and weekdays on which the system is accessed with timestamps.

The number of interactions of weekly usage was studied for each user over the chosen period, as illustrated in Figure 5.15.

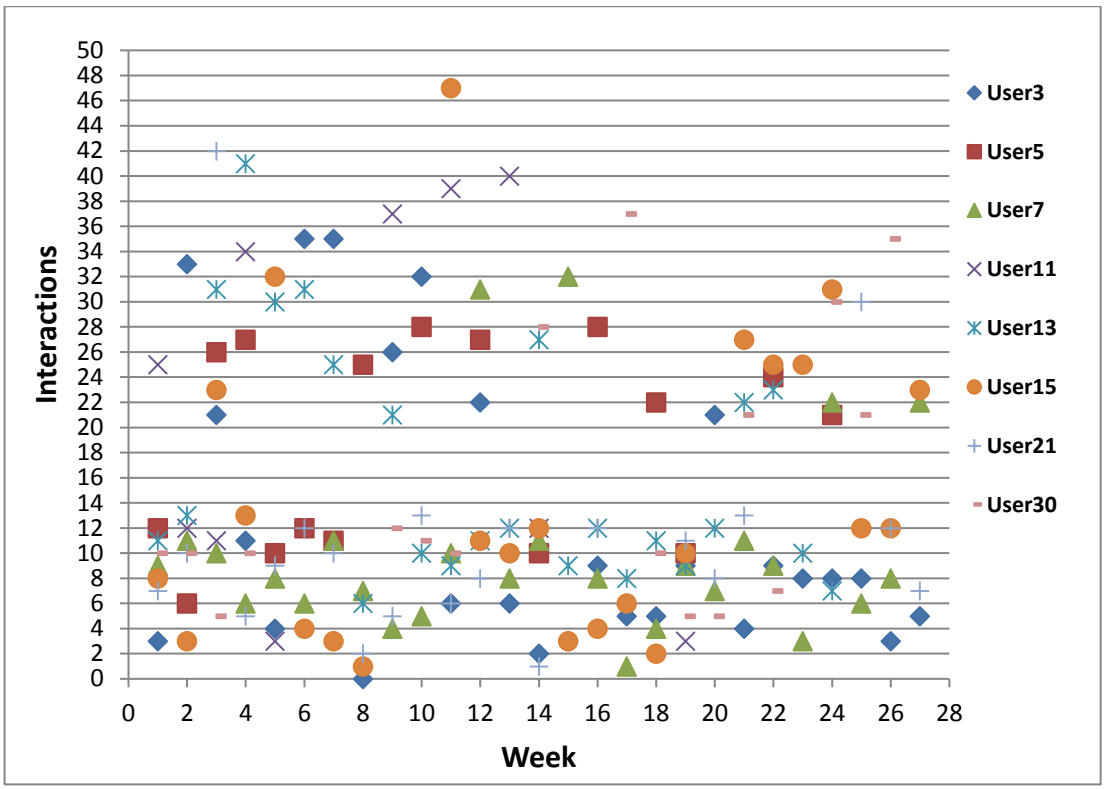


Figure 5.15: Total number of weekly users' interactions

Figure 5.15 shows the total number of weekly interactions of most users is between 2 to 12 and 20 to 36 interactions across the six months. Therefore, for any usage exceeds this expectation of interactions, the system could consider the current user as an intruder. However, some users, for some weeks, exceeded the determined range of interactions, such as users 11, 13, 15 and 21. This could affect negatively the rules and the system's decision because the system might bother legitimate users during their usage with an alarm, or stop, or reject them. However, other rules can be built to support the system's decision.

The number of file types of weekly usage can be examined for each user as another feature to increase the variety of rules that can help the system make an accurate decision, as demonstrated in Figure 5.16.

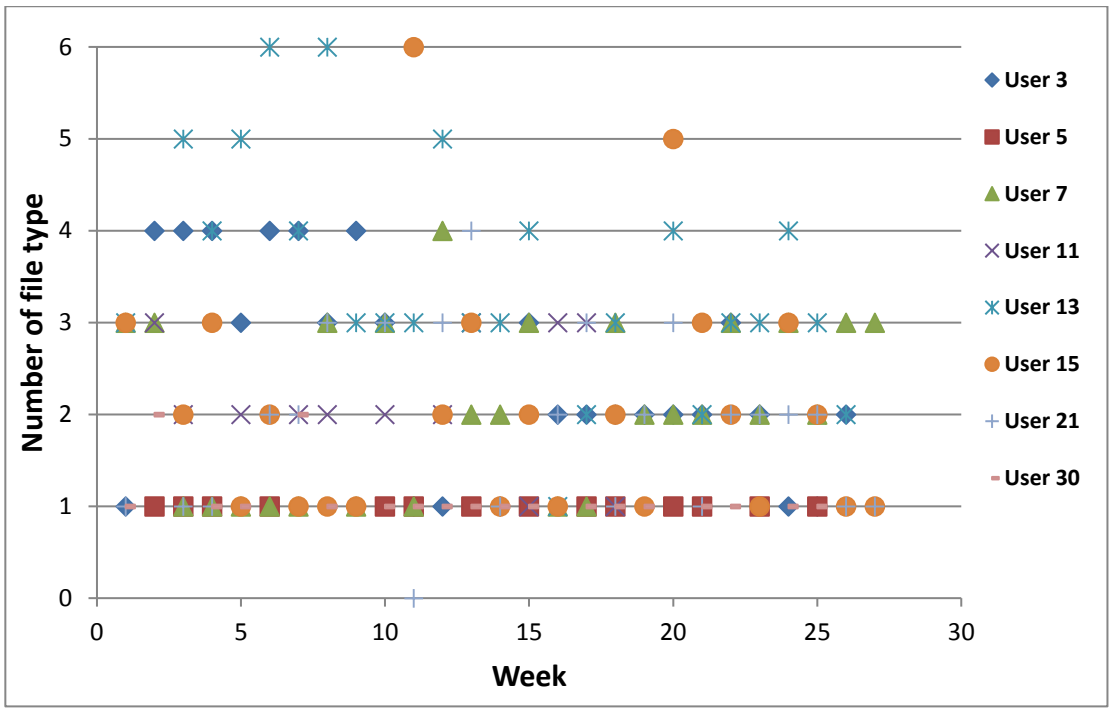


Figure 5.16: Total number of weekly usage for file types

The figure above shows the total number of file types for most users was between 1 and 4 a week across the six-month period. This can give insight into the system to make a decision when usage exceeds the determined system threshold. However, some weekly usage of some legitimate users exceeded the range, such as users 13 and 15. In this case, the system might make a negative decision for legitimate users.

Weekday usage could be also examined to add another feature for users of this group. Figure 5.17 shows that most users used Dropbox on weekdays except Monday and the weekend had a few users. Therefore, the system could consider the current user as an intruder who tries to use the services on Monday and on the weekend within a week. However, the system might make the wrong decision

with some legitimate users who work on these days, such as users 13, 15, and 21.

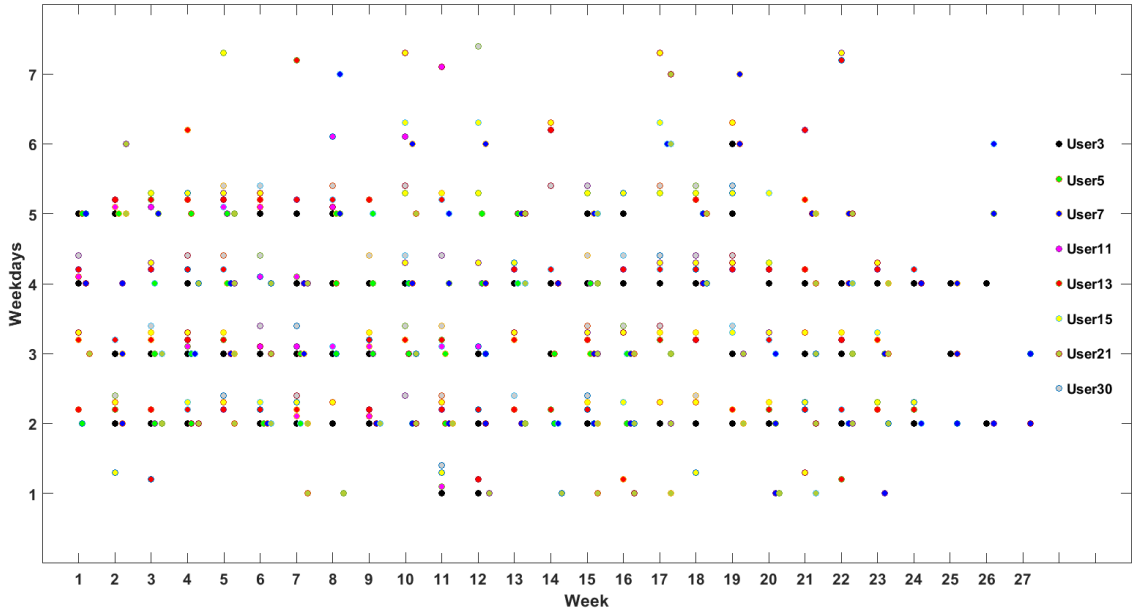


Figure 5.17: Users' usage during weekdays

Hourly usage could also provide another feature for this group to support the systems decision for protecting users from misuse. It was observed that most of users' usage/interactions occurred between 12 and 9 pm over the chosen time, as illustrated in Figure 5.18. Therefore, for a person who tries to use Dropbox out of the determined hours, the system could consider the current usage as a misuse. However, the figure also shows that some users used their account out of these hours, such as users 13, 15, and 21. Consequently, the system might make a negative decision with these users as illegitimate users.

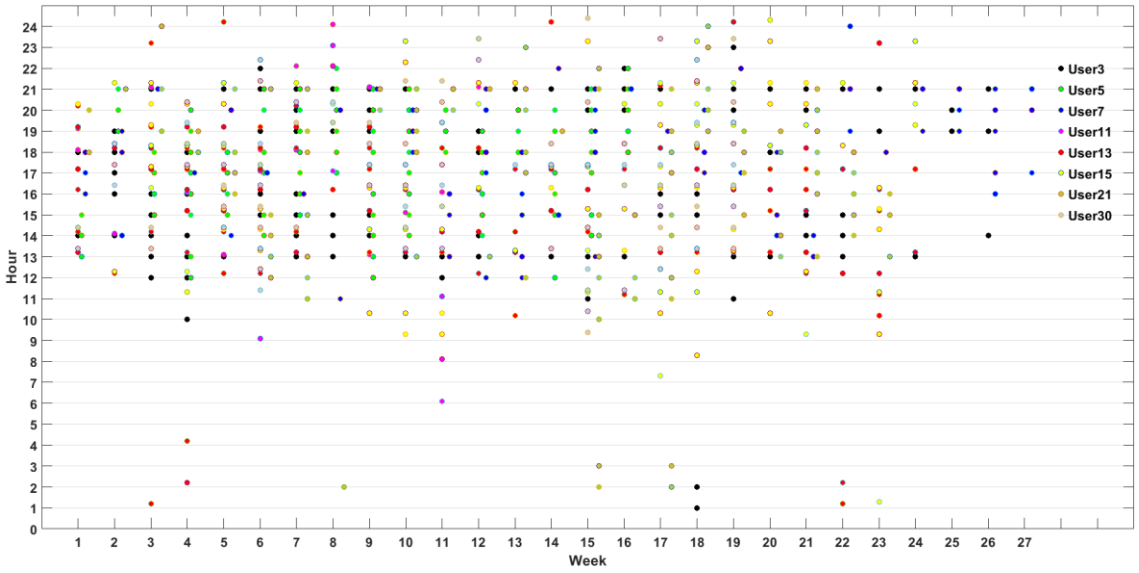


Figure 5.18: Hourly usage

Based on two-dimension visualisations of the above three groups, the descriptive statistical study identified several features from analysing the raw information presented by the available dataset (Dropbox activity). Some users had a smaller intra-class variance than others and had a large inter-class variance among others, which could provide good information to build strong user behaviour profiles to distinguish between users. For example, most usage of users 23 and 28 for the file types was different from other users' usage and they kept the same pattern of usage over the most selected period, as shown in Table 5.5. This can help determine the usage of these users from others easily. However, the intra and inter-class variance for some users was not ideal because of the inconsistent usage or the similarity of usage pattern among users such as Users 8 and 22 (as shown in Table 5.5). Analysing features in a single dimension does not give an efficient visualisation about the uniqueness of usage pattern of users, whereas multidimensional space can obtain a wide picture about the user usage by combining

multi-features together to provide more discrimination between users. Therefore, the next sections will apply a number of machine learning algorithms to examine the accuracy of discrimination between the users from multi-dimensional perspective.

5.3.2 Classification Algorithms

The descriptive statistical study identified a number of features that might provide rich information for a classifier. Therefore, this study will examine these features' effectiveness towards behavioural profiling from a multi-dimensional respective. As mentioned previously, four classification algorithms were applied in this study. The first classifier was SVM, which is based on a statistical learning technique. Two decision tree algorithms (RF and CART) were selected. The fourth method was the FF MLP neural network.

A more detailed analysis of the classifiers was undertaken to determine the impact optimisation would have. The results from the FF MLP neural network and RF methods are demonstrated in Figure 5.19 and Table 5.8 respectively. For the FF MLP neural network classifier, the best result of EER 6.98% was achieved by using 65 neurons while with the RF approach, the best performance of EER 9.93% was obtained when 25 trees were used. Neither SVM nor CART had any parameters to optimise.

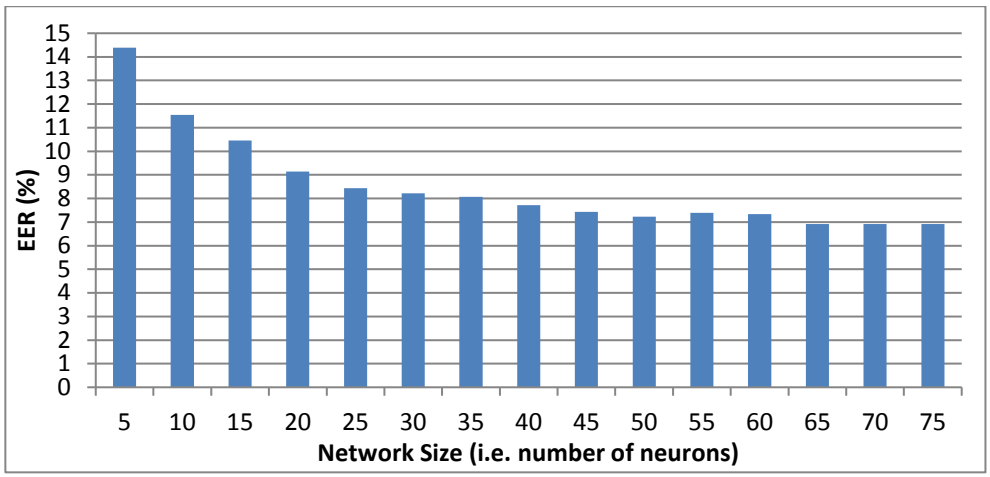


Figure 5.19: Performance of FF MLP with different network configurations

Table 5.7: Performance of RF with trees

Number of trees	EER (%)
5	10.39
10	10.41
15	10.21
20	10.18
25	9.93
30	10.26
35	10.43

The overall results of this experiment are presented in Table 5.7. Generally, the results support the idea of verifying the legitimate user or unauthorised access to data stored in cloud storage services, with EERs that are aligned to similar results in other applications from prior work.

Table 5.8: Performance of classification algorithms

Classifier	Time (D:H:M:S)	EER (%)
SVM	00:04:33:08	20.27
RF-25 trees	00:00:50:15	9.93
FF MLP Neural Network-65 neurons	02:02:40:55	6.98
CART	00:00:10:25	6.02

The nature of the classifier used does have an impact; however, with the exception of SVM, the variation in performance is not overly significant, which suggests

the classifier itself is not overly integral. As seen in Table 5.8, the CART was the fastest algorithm and achieved the highest accuracy with an EER of 6.02%. This would allow other factors, such as time taken to compute, computational overhead, and memory requirements to be considered as part of the selection.

Prior research has shown the volume of data per user has a significant impact on performance. As such, an analysis was performed where users were divided into two groups based on their interactions, as illustrated in Table 5.9.

Table 5.9: Users' performance with different classifiers

User	No. of Interactions	EER % based on classifier algorithms			
		SVM	RF	FF MLP	CART
1	549	19.08	8.88	10.32	5.13
2	585	2.11	3.47	1.95	1.96
3	652	6.23	9.68	4.65	3.35
4	677	2.52	1.56	1.43	2.02
5	726	26.80	6.55	7.21	3.63
6	764	23.79	8.17	13.09	5.71
7	797	3.02	28.29	4.78	14.89
8	1146	23.76	36.09	15.75	23.73
9	1370	10.67	2.36	4.70	1.22
10	1413	31.49	6.17	3.86	3.86
11	1462	13.47	10.08	4.89	5.36
12	1656	25.14	33.00	12.49	16.64
13	1714	11.25	0.11	0.22	0.02
14	1765	32.30	18.75	16.84	10.95
15	1988	39.35	19.86	15.02	15.56
Av*	**11501**	**18.06**	**12.87**	**7.81**	**7.6**
16	2250	30.22	4.69	7.81	4.08
17	2373	43.34	20.13	13.08	12.63
18	2487	28.53	19.87	9.13	12.27
19	2799	10.22	2.71	4.84	1.26
20	2879	17.54	0.56	2.40	0.54
21	2960	20.02	25.05	8.34	14.38
22	3226	28.17	1.33	3.57	0.99
23	3464	3.08	1.26	1.54	0.53
24	3568	31.55	3.95	13.88	1.91
25	4858	25.28	3.94	4.30	3.34
26	5780	7.13	0.15	0.19	0.15
27	6440	13.21	1.65	1.85	1.37
28	7263	29.06	11.20	7.78	5.89
29	14985	29.99	7.69	10.79	6.64
30	15013	19.81	0.68	2.79	0.75
Av**	**5356**	**22.48**	**7.00**	**6.15**	**4.45**

Av*: Average from the first group, Av**: Average from the second group

Table 5.9 shows that the first 15 users belong to the user group who have less interactions (i.e., equal to or less than 2,000 interactions) whereas the remaining 15 users belong to the user group with more interactions (i.e., more than 2,000 interactions). The selection of 2,000 was determined as sufficient to separate the groups yet ensure a suitable number of participants were left in each group. Based on the overall average performance from these two groups, it shows that users who have more interactions achieved better performance than users with less interactions by using the RF, FF MLP neural network, and CART classifiers. This can provide useful discriminative information to the classifiers that can help them to identify users.

However, this was not always true on a per user basis. For example, although users 17, 18, and 21 had more interactions than many other users, they achieved low performance than many users with low interactions, such as users 2, 4, 9, and 13. Further investigation suggests those three users used Dropbox as a backup solution by uploading photos, which could have been carried out automatically by a computer rather than the users themselves, creating difficulty for the classifiers to differentiate between user actions and computer-generated activities. This problem needs to be considered during the data collection phase to eliminate the automatic computer actions. Similarly, some low usage users got a better accuracy than many more active users. For instance, User 2 achieved less than 2% of EER across most approaches and User 13 got an EER closed to zero. When looking to the usage of these users, it was found that they worked constantly on specific file types that most of the other users did not use. This unique

pattern of usage made the classifiers easily distinguish the users. This result suggests that users who have more interactions achieve better performance in general. However, the uniqueness of interactions can be a key factor to build discriminative patterns for users, which can make classifiers more accurately distinguish between them.

5.3.3 Volume of Data for Training and Testing

As discussed previously in Chapter 4 (Section 4.6), the data splitting ratio for training and testing affects the performance of machine learning algorithms as learning by examples to find the pattern that distinguish the questioned classes.

This experiment studied the impact of the volume of data for training on performance. The CRT classifier was chosen for this experiment based on the outcome of the first experiment; also, the data splitting between training the classifier and testing the performance was set to 50/50, 66/34 and 80/20. Table 5.10 illustrates the performance of all users across the selected volumes of data.

Table 5.10: Performance based on volume of data with random selection

User	50/50	66/34	80/20	User	50/50	66/34	80/20
1	7.28	5.13	5.09	16	3.88	4.08	2.19
2	2.59	1.96	1.65	17	13.58	12.63	11.93
3	7.30	3.35	4.26	18	13.93	12.27	12.98
4	1.84	2.02	1.14	19	1.83	1.26	1.38
5	3.66	3.63	2.05	20	0.59	0.54	0.70
6	6.57	5.71	4.96	21	16.63	14.38	14.36
7	15.20	14.89	14.14	22	1.22	0.99	1.35
8	25.25	23.73	24.45	23	0.90	0.53	0.89
9	2.14	1.22	0.84	24	2.33	1.91	1.73
10	7.06	3.86	5.39	25	3.35	3.34	3.51
11	6.47	5.36	5.32	26	0.21	0.15	0.17
12	20.04	16.64	16.04	27	1.56	1.37	1.80
13	0.09	0.02	0.02	28	7.15	5.89	6.04
14	9.92	10.95	9.77	29	6.63	6.64	6.39
15	13.81	15.56	14.50	30	0.85	0.75	0.77
Average					**6.79**	**6.02**	**5.86**

As shown in Table 5.10, the training phase with a larger volume of samples achieves better performance than those with a smaller volume of data on average; the best result performance was 5.86% of EER achieved by using 80/20 splitting for training and testing, respectively. This is supported by the prior research as they suggest that larger volume of samples for training the classifier can have a positive impact on overall performance (Buschkes et al., 1998; Hilas and Sahalos, 2007; Yazji et al., 2009). This is logical as the classifier can be trained more about user behaviour patterns by using a larger volume of data, leading to a better performance. However, it is also worth highlighting that the change in performance from 6.79% EER to a best case of 5.86% EER is marginal. This suggests the nature of user behaviour across the six-month collection period is likely to be relatively stable.

From an individual user's perspective, the increasing volume of data for the training stage has different impacts on performance. When increasing the training data volume to 66/34 and 80/20 splitting, a number of users' performance improved and some stayed relatively stable. This suggests more data made little difference for those users. In a practical sense, being able to understand which users have more stable or active profiles would be useful in interpreting the classification decisions and in template retraining.

5.3.4 Time Series Sample Selection

In addition to the random sample used for the previous experiment (a standard methodological approach in feasibility studies), the impact of time and natural changes in user behaviour over time is important to evaluate. The data splitting

for training and testing of the CART classifier was the same manner as the previous experiment (i.e., 50/50, 64/34 and 80/20). The results of the experiment are presented in Table 5.11.

Table 5.11: Performance of the different volume of data with time series selection

User	50/50	66/34	80/20	User	50/50	66/34	80/20
1	18.38	15.86	15.15	16	2.22	1.37	1.13
2	2.61	3.84	4.28	17	34.52	32.90	35.51
3	6.49	6.22	6.41	18	25.87	19.43	15.45
4	2.92	1.18	0.04	19	2.10	1.61	1.15
5	18.28	25.05	21.06	20	8.54	4.51	5.32
6	2.43	9.85	3.54	21	25.10	22.64	19.03
7	19.90	17.24	19.45	22	1.74	2.30	1.89
8	41.39	40.57	43.59	23	1.90	0.96	0.82
9	2.36	3.15	2.80	24	6.27	5.34	4.57
10	27.60	38.91	23.38	25	4.92	5.09	6.32
11	10.07	8.89	4.71	26	0.11	0.19	0.31
12	40.31	38.13	36.66	27	3.28	1.91	2.29
13	15.69	19.24	14.40	28	11.28	11.05	9.74
14	17.32	17.05	19.60	29	8.77	8.35	10.22
15	27.37	21.14	23.83	30	0.81	0.83	0.85
Average					**13.02**	**12.83**	**11.78**

As demonstrated by Table 5.11, the best performance is EER 11.78% and it is achieved by using the 80/20 data splitting for training and testing. Similarly to experiment 2, the nature of the data split has not had a significant impact on performance; however, the results themselves have doubled. This suggests that over time, user behaviour changes and, therefore, care must be taken on ensuring appropriate template renewal procedures are developed to maintain levels of performance.

5.4 Discussion

The main goal of this chapter was to examine the ability of creating user behaviour profiles from users' interactions during interacting with cloud storage service (Dropbox). Then these profiles employed to identify the abnormal usage of users that deviates from the normal user behaviour patterns. As a result, a number of

experiments were applied on the dataset that was collected from Dropbox to examine the validation of this goal.

The first experiment implemented the statistical analysis. The main purpose of this method was to analyse the available features that could contribute to building a sufficient user behaviour profiles. Inter and intra-classes variance for these features was examined to each user. Many users had valuable discriminatory information, which could be helpful for identifying many misuse scenarios. For example, using the service outside the usual time of usage can be identified or uploading/deleting different file types that have not been used by a user. However, some users showed that their information usage was fairly poor to identify because of the limited number of interactions or the features derived from the available dataset for these users were similar to other users. Therefore, an intelligent system needs to be developed to determine users who can implement the behavioural profiling successfully.

The results of the second experiment reveal that cloud storage service users can be discriminated via their usage with a reasonable performance being achieved. In addition, the outcome of this research is in line with the highest results achieved in the related works such as Shi et al. (2011), Aupy and Clarke (2005), Yazji et al. (2014), and Subudhi and Panigrahi (2015). Concerning the performance of each individual classifier, the CRT algorithm achieves 6.02% EER and outperforms the other three chosen classifiers (i.e., SVM, RF and FF MLP neural network). From an individual user's perspective, on average, users who have more frequent activities/interactions acquired better results than those who had fewer interactions

across most classifications. However, users with fewer interactions also achieved a good level of performance. For example, when examining the interactions of those users (e.g., users 4, 9, and 13), they had a unique way of using Dropbox (particularly unique file types). Therefore, a good pattern (uniqueness) from the users' interactions can also affect the performance of the classifiers even though the number of users' activities is low.

The results of the third and fourth experiments show that the data split for training and testing the classifier and the timestamp factors have an impact on the overall performance. As shown in these two experiments, a larger volume training data (i.e., 80/20 splitting) with random sample selection achieves better performance with 5.8% of EER on average. However, regarding individual users, the performance of a number of users with more training data (i.e., 66/34 and 80/20) is not as good as the results being achieved by using less training data (i.e., 50/50 split for training and testing). One of the reasons could be that part of the dataset was collected from early-stage PhD research students, and normally they conduct various activities and use different file types during the initial research period. Therefore, they might deal with specific files types and actions within the first period of their research, then other file types and actions with the next period. These changes in user behaviour can affect the classifiers' performance because their activities are so diverse.

When applying the behaviour profiling technique in practice, the time series sample selection showed a significant difference over the random sampling. Therefore, user templates need to be updated regularly to ensure their quality for

achieving a high level of system performance. However, the renewal of users' templates dynamically is not an easy task because it might need to avoid including impostor's behaviour with the legitimate behaviour. For example, an impostor might be accepted by the system over time as the genuine user as more and more impostor samples are included within the template renewal process. This problem needs to be managed carefully and correctly to avoid capturing illegitimate usage while ensuring users' convenience level exists in the system for legitimate user comparison for sample selection techniques.

5.5 Conclusion

The results have successfully demonstrated the ability to correctly discriminate between users based on their interactions derived from cloud storage (Dropbox). Accurate user behaviour profiles can be built to help distinguish between the normal and abnormal usage. Classification algorithm experiments achieved high accuracy with only the SVM not performing particularly well. Further experiments have shown that time-series versus random sampling of data for training does have a significant impact on performance; however, this is less so for the volume of training data. From an individual's performance, many of participants achieved a high performance where the system was capable of identifying their interactions fully correct without any error. Subsequently, the approach proved a highly promising solution to applying user behavioural profiling as a supporting technique to validate the users after initial point-of-entry authentication. This can contribute and guide the system to identify a misuse of cloud services in continuously and friendly manner.

However, there were a number of users who performed particularly poorly and in line with most behavioural-based applications, would not be suited to such a technique. Moreover, there is concern about the performance of the available dataset when moving forward in practical use (real world) based scenario in terms of speed detection with volume of data that are needed to get a reliable outcome, as well as using time series rather than bootstrap method for data selection. In addition, there is the aspect of when and how template generation or renewal should be undertaken. These aspects need carefully consideration to make sure the performance level is optimally achieved. To examine the feasibility of applying behavioural profiling with another level of cloud service, the next chapter will present a study that collects user interactions with another cloud service (IaaS). The nature of interactions can be different and of the volume of these interactions might be more as the user can interact with various applications. This can help to extract various features that might contribute to good performance and allow for implementation of other investigations.

6 User Behavioural Profiling Infrastructure as a Service

6.1 Introduction

The previous chapter showed the feasibility of applying behavioural profiling with the top level of cloud computing service (software as a service) as a second factor technique to verify the subscribers after the initial point-of-entry authentication. This chapter attempts to investigate the same technique (behavioural profiling) with IaaS, as it is a vital layer that supports all cloud services including SaaS and PaaS. Importantly, as this layer is the underlying layer for all top layers, it might give an opportunity to collect more data because users can build and interact with various services not only single service like Dropbox. This can help to deal with some issues that occurred with the previous dataset. One of the main issues was lack of user interaction.This factor can have a vital impact on system performance, speed of detection, and creation or renewal of accurate users' templates. Therefore, more user interactions can help to extract unique patterns for the users. These patterns can be used to build user behaviour profiles, which would be subsequently applied to continuously verify the identity of the current user and detect a misuse in the service.

To thoroughly study the possibility of applying behaviour profiling techniques on cloud infrastructure service users, this chapter seeks to collect a real dataset from these users to be implemented in a number of investigations. These investigations include applying descriptive statistics to analyse the dataset and know its nature more deeply. This can help to explore the discriminative features that are

available in this dataset to build reliable user profiles. Implementing machine learning algorithms with different strategies on these profiles to explore how they are reliable to be implemented in reality in terms the performance will then be discussed, including a comprehensive discussion on the feasibility of this approach to increase the security of cloud infrastructure services.

6.2 Methodology

6.2.1 Data Collection

To achieve the statistical and practical experiments, a real dataset of users' interactions with the cloud infrastructure application needs to be collected. However, to the best of author's knowledge, there are no public datasets that would be used for this study. Moreover, the collection of users' activity in cloud computing services has proven to be problematic because cloud providers are unwilling to provide such access directly because of privacy and security concerns. Whilst it is possible to create IaaS-based images and have a population of participants use these machines for a specified period, it was felt this might result in behavioural patterns that do not truly reflect user's normal activities. Consequently, a decision was made to capture users' interactions on their own personal computers (applications and websites) simulating the environment of a cloud infrastructure service. To collect these activities, software was created and installed on the participants' computers. As mentioned in the previous chapter, regarding the number of participants and the environment of collecting data, the same environment was selected but the number of participants was doubled. The increase in the number of participants was to explore the impact on system performance and to determine

the degree to which the proposed technique was reliable for dealing with this increase in number. Therefore, the activities of 60 participants (including PhD researchers and undergraduate students) were obtained by instilling the software during a three-week period (02/09/2017 to 23/09/2017) on their computer desktops, resulting in a private dataset containing 1,048,195 user interactions. The interactions comprise the following information: the start and the end time of applications being used (e.g., Excel, Word and MatLab) and web services (URLs) being visited. The data was anonymised to protect the participants' privacy and ethical approval was sought and obtained from the authors' institution (Appendix A). Table 6.1 demonstrates a sample of user actions within the dataset.

Table 6.1: User activity with personal computer

Day	Hour	Minute	Second	App/URL	Event
2	9	10	8	Word	Focus
2	9	21	16	Word	Lost focus
2	9	21	20	Endnote	Focus
2	9	23	44	Endnote	Lost focus
2	10	15	30	Paint	Focus
2	10	45	23	Paint	Lost focus
2	10	45	30	v2wLG+IIc…	Focus
2	10	49	13	v2wLG+IIc…	Lost focus
2	11	17	55	Ri9SK2bSH…	Focus
2	11	19	34	Ri9SK2bSH…	Lost focus

Table 6.1 shows there are different ways to extract unique features for users, which can help to recognise their usage from other people. For example, as part of the dataset is collected from PhD researchers, some of them of early stage

might mostly visit different websites searching for different articles to read. As researchers in their final year might mainly focus on writing their thesis, they might deal with Microsoft Office, such as Word and Excel applications. Moreover, some mid-stage researchers might commonly focus on the practical part of their work which might deal with a specific application to achieve/analyse the results. The time, duration, and date of the aspect of access might also be another factor for distinguishing the users from others. For example, some students come to university and use the services during the working hours' days. Therefore, when an impostor accesses the account service of another user, he/she may use the services differently by opening different applications and visiting different websites that the owner might not use, or the date and time of accessing the service might be different, such as accessing the service at midnight which the genuine user might not usually do. As a result, a variety of behavioural patterns can be observed while users interact with services. These patterns can help discriminate between legal and illegal usage.

As seen in the previous experiment of Chapter 5, the number and nature of users' interactions are vital factors to generate good user behaviour profiles through giving a better understanding of users' usage pattern. Therefore, the volume of these interactions and their nature need to be studied for the current dataset of the selected users. Table 6.2 below illustrates the data that was gathered from 60 participants during a three-week period of usage (totalling 1,048,195) on their computer desktops.

Table 6.2: Summary of the dataset

Total number of unique App/URL	6,710 (0.64%)
Total users' interaction of shared App/URL	924,429 (88.19%)
Total users' interaction of unique App/URL	123,766(11.81%)

It can be noticed from the above table that the dataset contains a rich information including applications and websites having the total number (1,048,195 logs). The total number of users' logs shows that it could be enough to build user behaviour profiles. However, there is a need to know the distribution of these interactions for each user based on a number of days. Additionally, each user's usage needs to be studied more deeply and compared with other users to see if there is any possibility to distinguish the users from each other, as will be shown later in the section of descriptive statistics.

6.2.2 Procedure

The main aim of this study is to focus on understanding the degree to which behaviour profiling can be successfully applied to verify users via their usage within cloud infrastructure services—understanding whether it is the genuine user or not to provide a basis for a security system to respond. Therefore, a series of experiments were conducted on a real dataset to examine the impact of a number of factors on the performance of the machine learning algorithms. These include two further research questions:

- Does the volume of data and sampling selection for training and testing have an impact on performance and classification algorithms?

- How much data and time are required to generate a user template within specific criteria?

Three experiments were developed to investigate whether the collected data can be used to differentiate among those users whom created this data.

The first experiment will be a descriptive statistical approach. This includes exploring the nature of the dataset from different aspects such as the volume of data, type, and the period of usage for each user. This can help to examine the possibility of extracting unique patterns from the selected features to discriminate individuals to build sufficient user behaviour profiles. Then, machine learning algorithms were applied on these profiles to explore the degree of the reliability in reality.

To examine the reliability of the selected technique in practice, the second experiment implemented the best two classification algorithms (that achieved a better performance with the SaaS study) on the given dataset. The configuration of the two selected classifiers remained as default based on prior studies. Three splitting approaches for training and testing data were applied on these algorithms: 50/50, 66/34 and 80/20 to examine the impact of data volume on the overall performance. For each data volume setting, two sample selection methods were used: a random sample selection across the dataset and a time series sample selection (i.e., samples are selected sequentially as in a reality sense). This can help to understand how this approach can perform in practical sense. Additionally, the outcome of this experiment would explore how the performance of the system is affected by investigating the nature of different classification approaches.

Therefore, the optimal classifier can also be identified based on the findings of this experiment. Finally, the comparison between the accuracy of the result of each data volume would give a better understanding of the nature of user behaviour profiles with the impact of the sample selection on the performance of the algorithms.

The third experiment focused on exploring how much training data is required to generate a user template with an acceptable level of performance. For the purposes of this study, an EER of 10% was set, as the average performance that can get from behavioural profiling techniques based on prior studies. In practice, a user profile would need to be created based on time-series rather than random sample selection. This is because when the system is applied in reality, the sample will be entered sequentially to the system. As such, time-series sample selection was applied to achieve the goal of this experiment; the first day's data was used for training and the data from remaining days was employed for testing, then the data from the first and second days were used for training and the remaining data from the rest days was used for testing and so on, as shown in Figure 6.1.

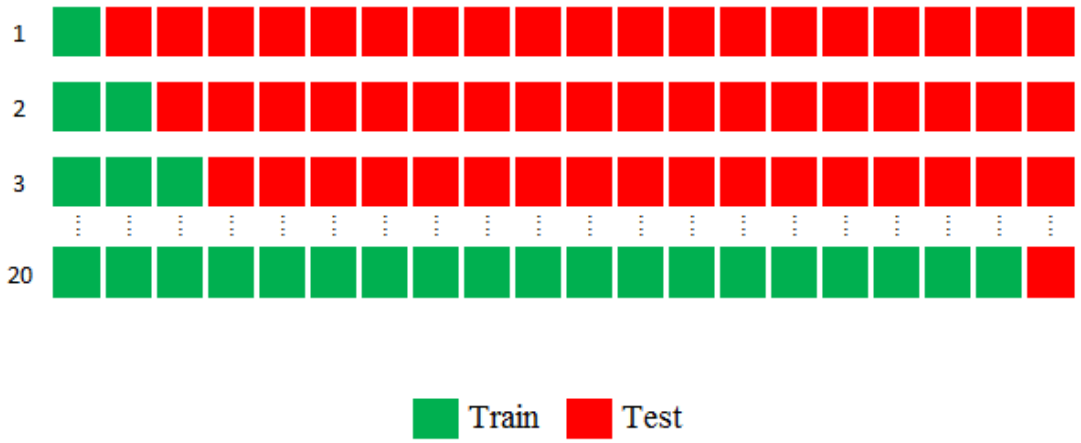


Figure 6.1: Taring and testing procedure

To achieve the practical experiments of classification algorithms, four features set were used in this study which are the hour, minute with applications and URLs. These features could provide a good level of pattern recognition amoung the users. To make those features acceptable by classification algorithms, the symbolic-valued attributes (e.g., the name of applications and URLs) were enumerated into numerical attributes and into the range of 0-1 (Sola and Sevilla 1997), as have been done in the procedure of the previous experiment in Chapter 5.

In addition, the same previous method was applied to divide the dataset for training and testing the classification algorithms with a similar procedure to compute the accuracy of each user. Additionally, as the given dataset is larger than the previous dataset which needs more time to be achieved, two faster and well perform accuracy classifiers were selected from the four classifiers which have been used in the previous experiment in Chapter 5, which are CART and RF.

6.3 The Experimental Results

This section seeks to first analyse the gathered data by applying a number of descriptive methods, followed by analysing and discussing the results of the classification algorithms.

6.3.1 Descriptive Statistics

A number of statistical methods were examined on the following features: the time, applications being accessed, and websites being visited. The volume of interactions, mean and standard deviations, and median with first and third quartiles are examples of these statistical approaches that were applied. These examinations could give an insight into the relationships that occur within the data. It also shows the possibility of establishing the degree to which input samples of data are similar or dissimilar. Therefore, this section seeks to find the similarity patterns of samples of each individual user and how these patterns are different from others.

As shown in Table 6.2, the total number of logs of users could be considered rich information that can be investigated to enable the study moving forward for meaningful analysis. However, interaction's volume of each user needs to be enough to generate these profiles. Therefore, the first attempt will investigate the volume of interactions with the number of days of usage for each user over the given period, which could also help to separate the usage of each user from other. Figure 6.2 demonstrates the total logs and days of each user.

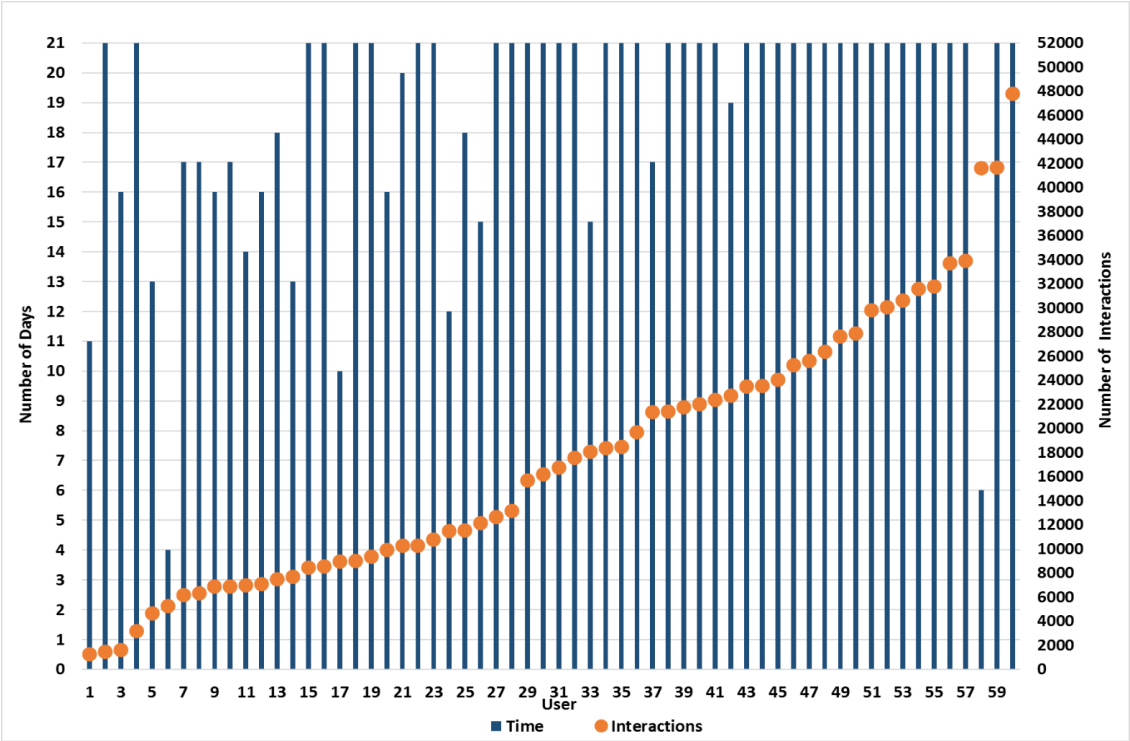


Figure 6.2: Total Users' interactions of the dataset over 21 days

Figure 6.2 shows the significant diversity of usage interactions and days between users, which could be offered an opportunity to provide a discriminative feature that can distinguish among the users. This diversity in the number of users' interaction can also be used to distinguish between users. For example, the first three users 1, 2 and 3 can be separated easily from the last three users 58, 59, and 60 based on the number of interactions. However, the figure also shows that some users had approximately the same number of interactions such as users 9, 10, 11, and 12. More importantly, some users accessed the service from time to time, as the dataset is collected from students, they might be part-time students. For example, user 6 come only one day a week and user 58 come two days a week, so they did not use the service that often. While the most other users use

the services across all given weekdays. This feature can be investigated to distinguish the usage among users. However, this is still not a strong factor for distinguishing the users because the similarity of usage within the day itself might be a big problem if the usage patterns of these users are similar. Therefore, deeper investigation is needed to understand the nature of these patterns for each user.

The number of applications and URLs was calculated for each user to see the volume of each of them for each user, as illustrated in Figure 6.3

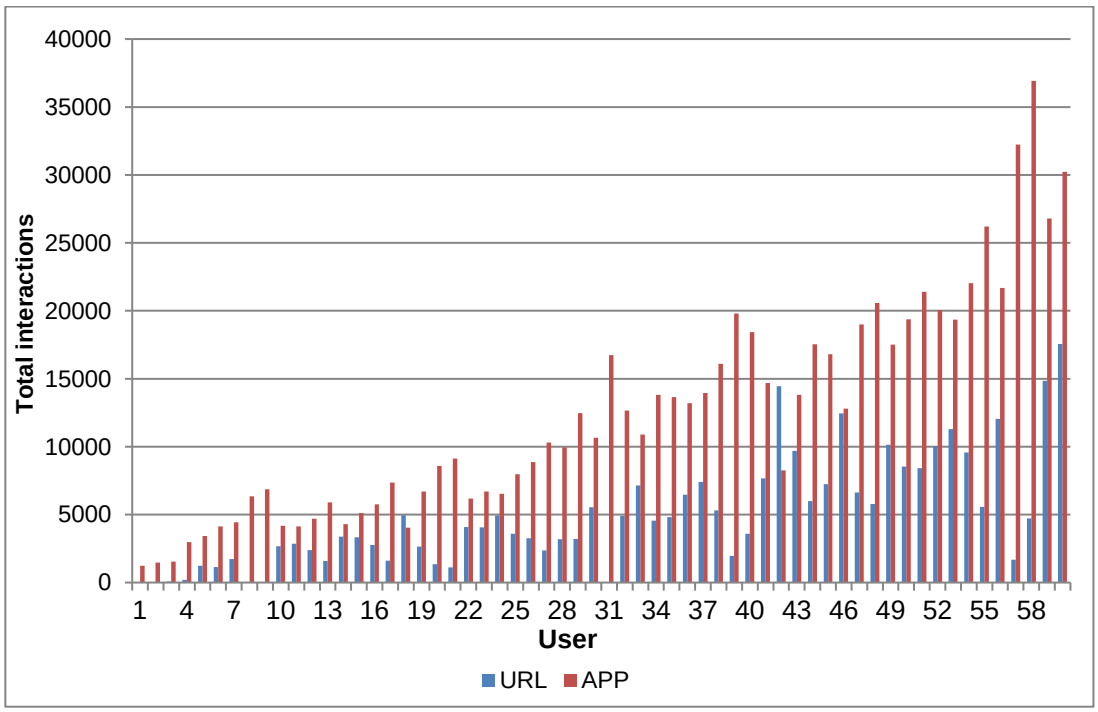


Figure 6.3: Users' volume interactions for Apps and URLS

Firstly, Figure 6.3 shows the possibility of creating behavioural profiling for most user-based application usage or websites being accessed. Therefore, multi-behavioural biometrics profiles can be generated for each user except some users

such as users 1 and 2, as they did not access any website during the given period because they used a virtual machine to implement their work. Additionally, as users 1 and 2 did not have access to URLs during their usage, they can be distinguished from other users who access websites frequently, such as users 42, 59 and 60. However, some users have roughly a similar usage of applications and websites, such as users 8 and 9 or 10 and 11. Moreover, distinguishing among users based on total interactions of all given period might not be enough because if the system is designed to make a decision based on hourly or daily usage, the users might have similar volume of usage for these two metrics within a specific time window. Therefore, a unique pattern across all interactions needs to be discovered.

The percentage of the total number of unique applications and URLs was computed, as shown in Figure 6.4

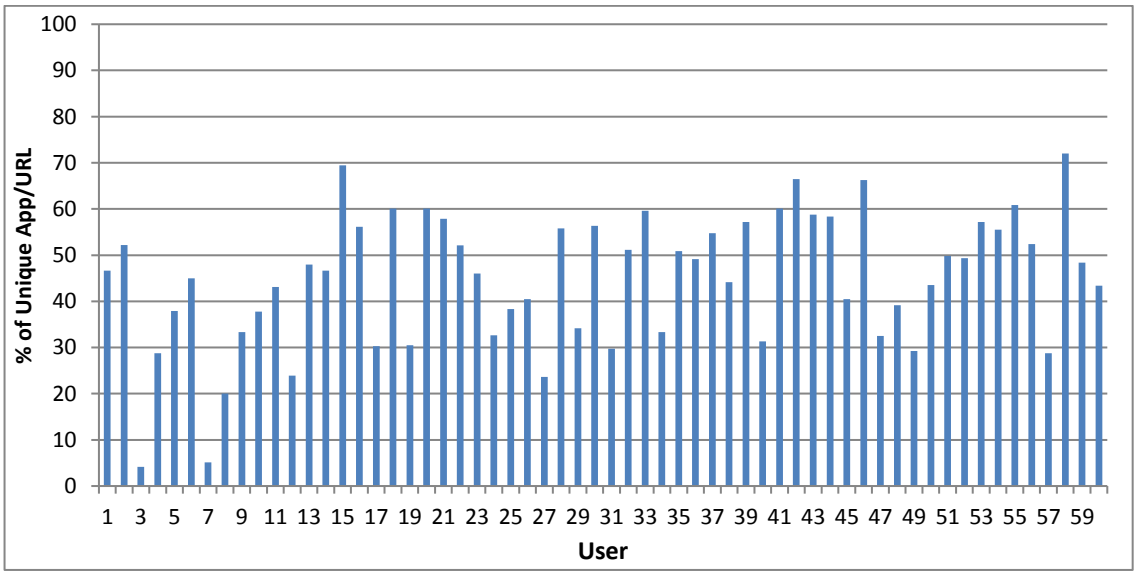


Figure 6.4: Percentage of the total number of unique Apps and URLs

Figure 6.4 shows that the percentage of the total number for unique applications and URLs can be considered too high for the most of the users; half of the users have nearly 50% of their usage being unique for the name of the application and the URL. This high percentage could strongly support the distinguishing process for the users. However, when looking at the volume of interactions for these unique applications and URLs, it is different, as demonstrated in Figure 6.5. This because the user might be open/access an application/URL only one or two time within a long period of time, but he/she might open/access some popular applications/URLs more often, such as Microsoft Office applications and university websites.

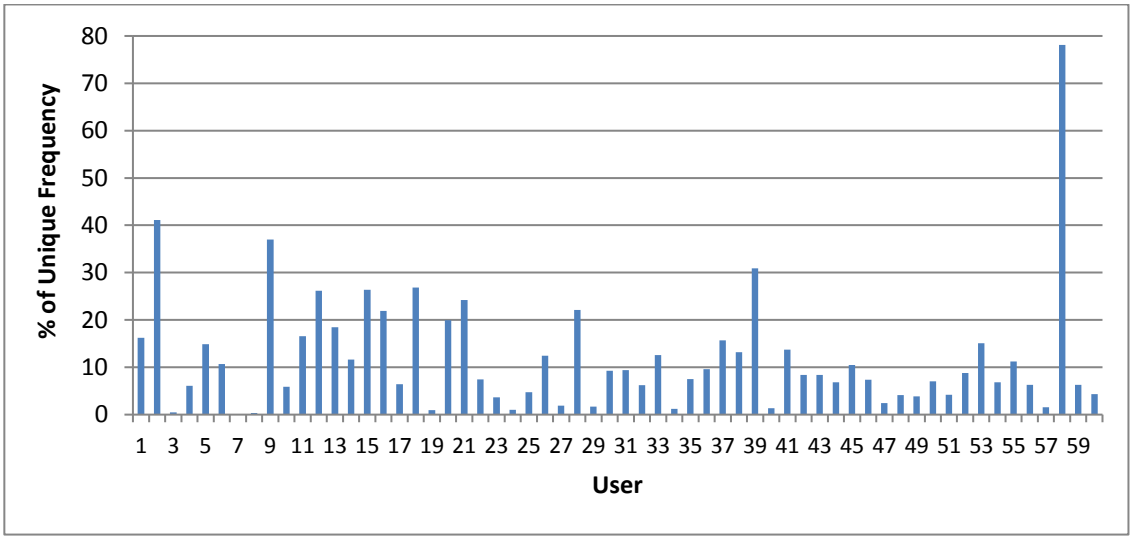


Figure 6.5: Total Volume of Unique Apps/URLs

Figure 6.5 still shows that some users have a unique usage for the application and websites more than 30% of their usage such as Users 2, 9, 39 and 58. This unique usage could help a classifier to recognise these users from others easily.

However, the figure also shows that the most users have mainly shared the applications/URLs usage with other users which could cause insufficient user behaviour profiles to be built. Therefore, other features need to be investigated to examine whether there is any other possibility of solving the problem of sharing same applications/URLs. Mean and standard deviations were applied to the daily usage of the users, which could give a possibility to understanding the degree of the similar or dissimilar in users' usage, as illustrated below in Figure 6.6.

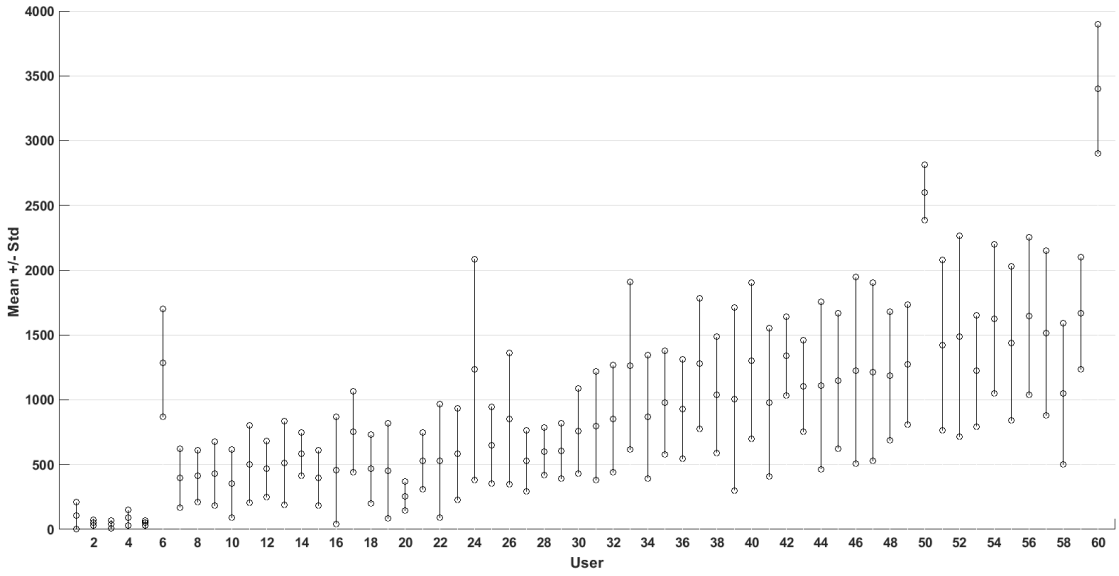


Figure 6.6: User mean & standard deviation of daily interactions

Figure 6.6 shows that some users have different average of daily usage, particularly users 50 and 60, who did not overlap with other users. This can have a positive effect on identifying users. However, with a fairly large spread of the standard deviations of most users, which made their usage overlap, this can have a negative impact on the performance of classifiers because of the inability to discriminate between users.

More analysis has been done by selecting one application and one URL which have the largest volume of interactions across all others and then computing the mean and standard deviation across all users of daily usage, as shown in Figure 6.7 and Figure 6.8.

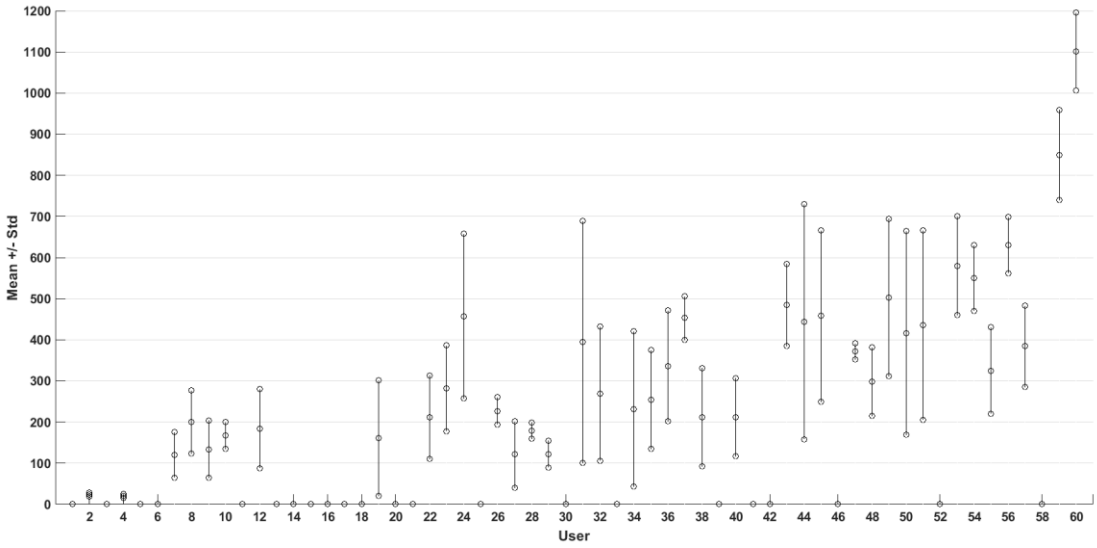


Figure 6.7: User mean & standard deviation of daily usage for the largest application

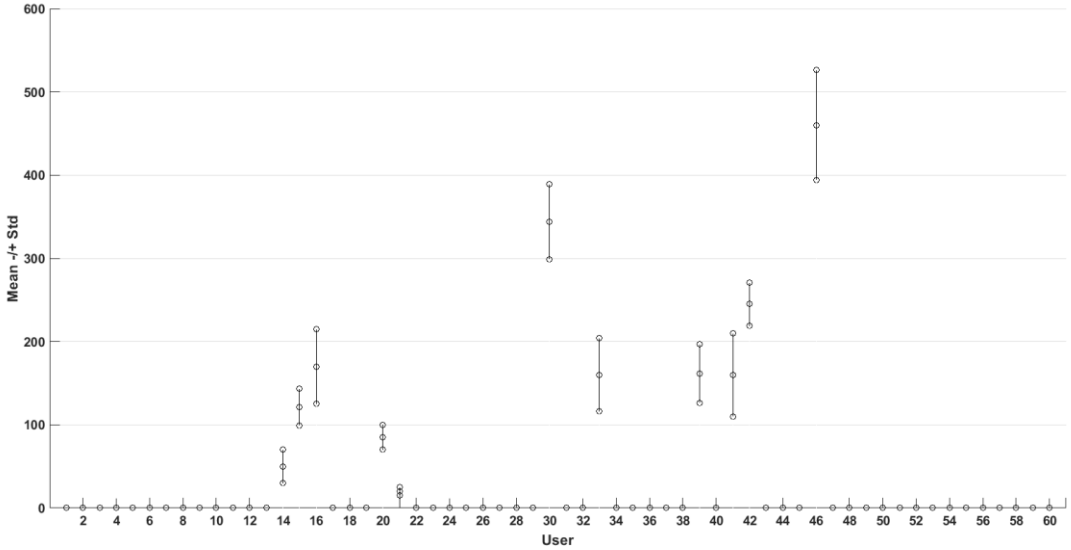


Figure 6.8: User mean & standard deviation of daily usage for the largest URL

From the above two figures, it can be noticed that some users can be distinguished based on the selected application and URL. For example, users 59 and 60 in Figure 6.7 had a different average of usage for the selected application, which is not overlapped with other usage of users. Similarly, users 14, 20, 21, 31, 43 and 46 in Figure 6.8 had also a different average of usage for the selected URL. Therefore, the usage of these users could be easily identified. However, most other users, especially users in Figure 6.7, had overlapped in average usage, which could make classifiers unable to distinguish among users.

Finally, minimum, maximum, first quartile, median, and the third quartile were applied to see the distribution of hourly usage across all users, as illustrated below in Figure 6.9.

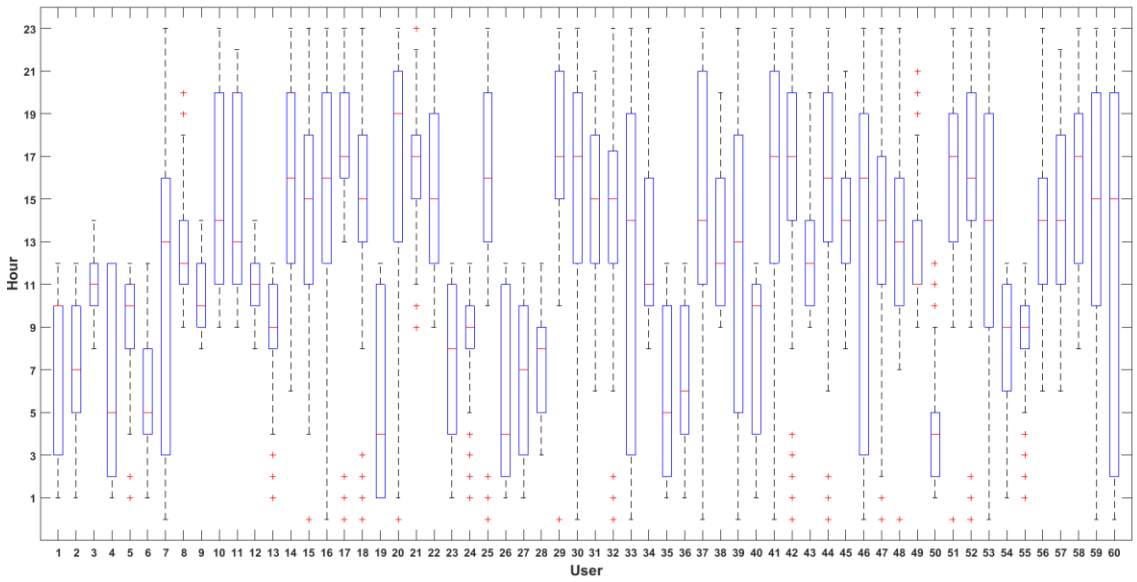


Figure 6.9: Distribution of users' based on hourly access

Although Figure 6.9 shows that some users have a similarity usage of hourly access for applications and websites, such as users 10, 11, 14 and 16. This can

make the recognition process of the users' usage more difficult. However, other users had different access of usage. For example, users 24 and 50 work only in the mornings across the given period, whereas users 17, 21 work only during the afternoon. Therefore, it can be verified between these users easily based on the access time.

6.3.2 Various Train/Test Set Ratio with Two Sampling Methods

This experiment studied the impact of the volume of data for training with time series and random sample selection upon the performance. The CART classifier was chosen for this experiment based on the best outcome of the first experiment; also the data splitting between training the classifier and testing the performance was set to 50/50, 66/34 and 80/20. Table 6.3 illustrates the overall performance of all users across the selected volumes of data with the two sampling methods.

Table 6.3: Performance of classification algorithms

EER (%) based on the volume of data						
	Time-series selection			Random selection		
Classifier	50/50	66/34	80/20	50/50	66/34	80/20
RF	15.35	13.18	12.40	4.07	3.57	3.09
CART	8.51	7.35	6.55	0.69	0.44	0.32

The results of this experiment (as illustrated in Table 6.3) are encouraging to support the idea of verifying the genuine user or identifying misuse of unauthorised access to cloud infrastructure services. The table also shows that the nature of the classifier had a significant impact on improving the system performance. The CART algorithm achieved a higher accuracy than the RF method regardless the amount of data being allocated to training and testing. This includes the time series and random sample selections with the highest accuracy of 0.32% EER.

From the sample selection perspective, Table 6.3 shows that the random sample selection achieved better performance than the time series selection within both classifiers and across all volumes of training and testing data split. This can be attributed to the high probability of selecting various user activities across the entire usage range whilst employing the random sample selection. It is also worth highlighting that the change in performance with both types of sample selection (random and time series) gets better as the amount of training data increases; decreases of 2% and 0.37% in EERs can be observed for time series and random selection respectively. This suggests that the nature of user behaviour across the three-week collection period is likely to be relatively changeable. Therefore, care must be taken to ensure appropriate template renewal procedures are developed to maintain performance levels.

The classifiers' overall average performance in terms of data volume (as illustrated in Table 6.3) also shows that the training phase with large sample volumes achieved better performance than those with smaller data volumes. Based upon the overall average individual performance using the CART classifier with 80/20 of data splitting and random sample selection, the trend line regression approach, as illustrated in Figure 6.10, also supports the same idea. Users with high volumes of interaction achieved better performance than users with fewer interactions. This support the idea put forth by prior research that more volume would provide better accuracy (Buschkes et al., 1998; Hilas and Sahalos, 2007; Yazji et al., 2009). Additionally, it is logical as the classifier can learn more about the pattern usage of a user by acquiring a large volume of data, leading to better performance.

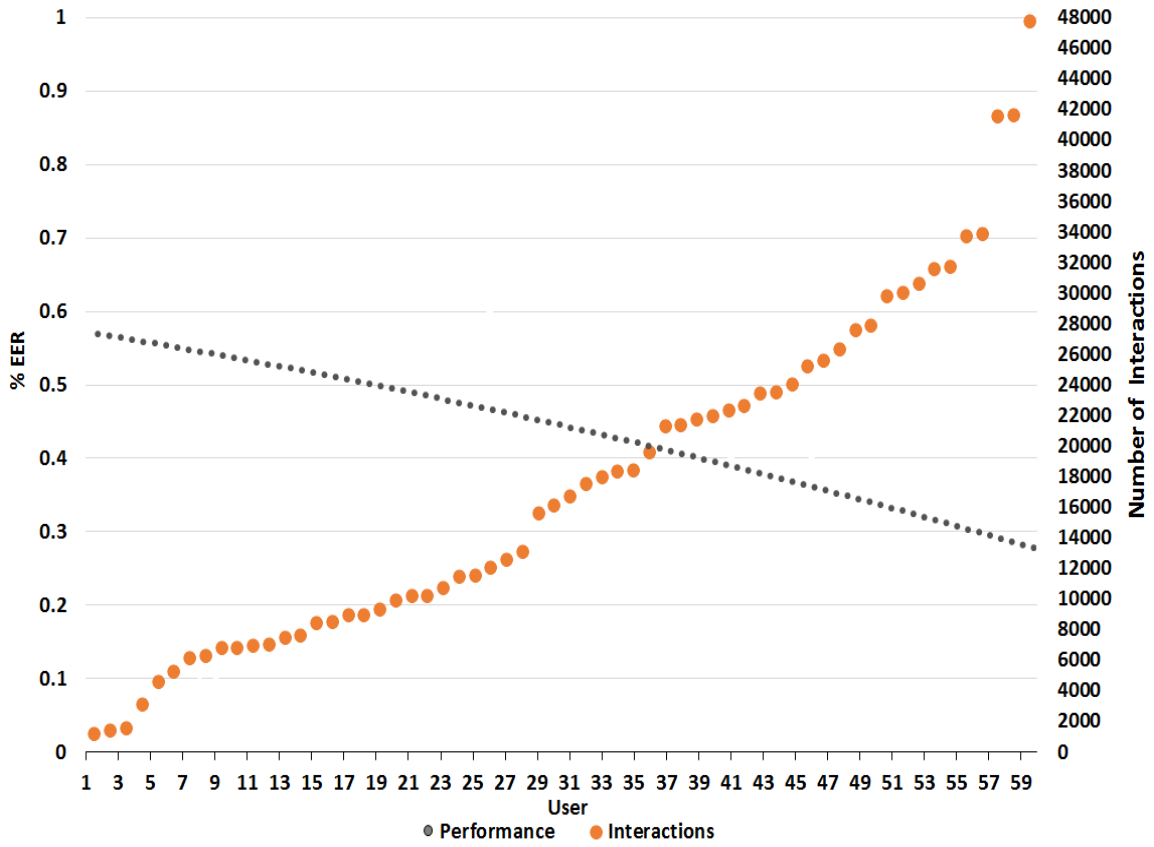


Figure 6.10: Average performance based on volume of data

6.3.3 Time and Volume of Data Required for Generating Users Templates

This experiment focused on the amount of data and time required for each user to generate a user template based on predefined criteria (10% of EER). The CART classifier was chosen for this experiment because of its outstanding performance in the first experiment. In a practical sense, the data split between training the classifier and testing the performance was selected based on using the daily basis as a time window, as mentioned previously. Figure 6.11 demonstrates the statistical distribution (min, median, and max) of the performance of all users across 20 days.

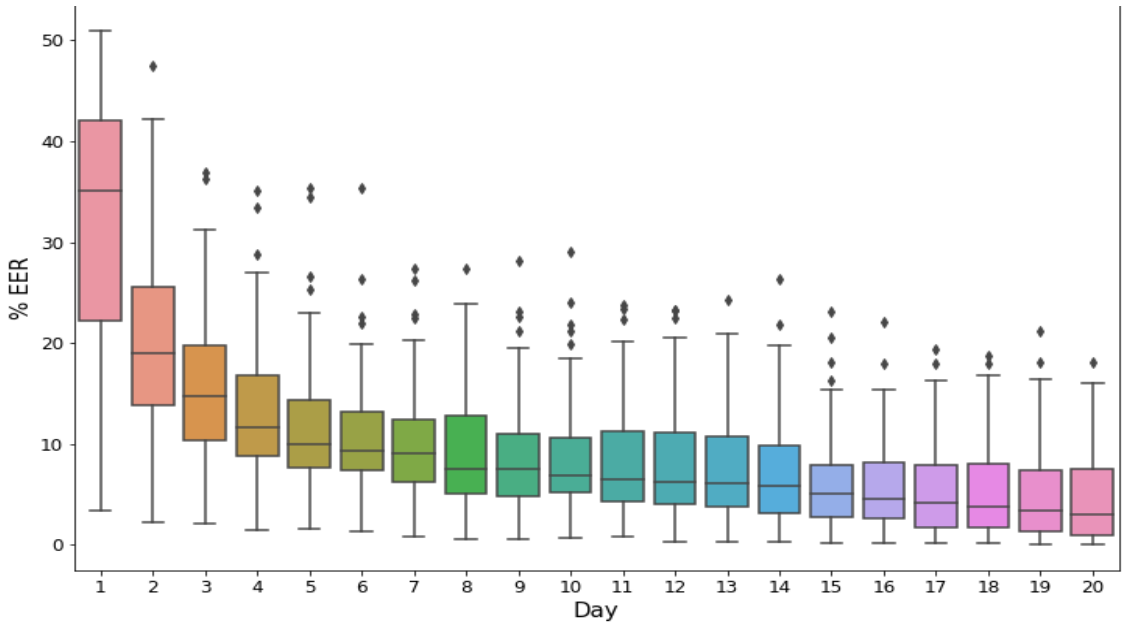


Figure 6.11: Distribution of users' performance across 20 days

Figure 6.11 shows the classifier achieved a significantly higher performance for larger volume of training data than the low volume of samples for the training phase, specifically within the first five days. Therefore, based on the overall distribution of users' result accuracy, it suggests that at least five days of user data are needed as an overall average time to profile individuals within the given criteria. However, it can be seen on the chart that there is also a variation among actual users; some users would need less than five days and others would need more to generate the template. Therefore, further investigation is required to determine the actual time and interactions required for each user. Figure 6.12 demonstrates the minimum days and interactions needed for each user to build suitable user behaviour profiles.

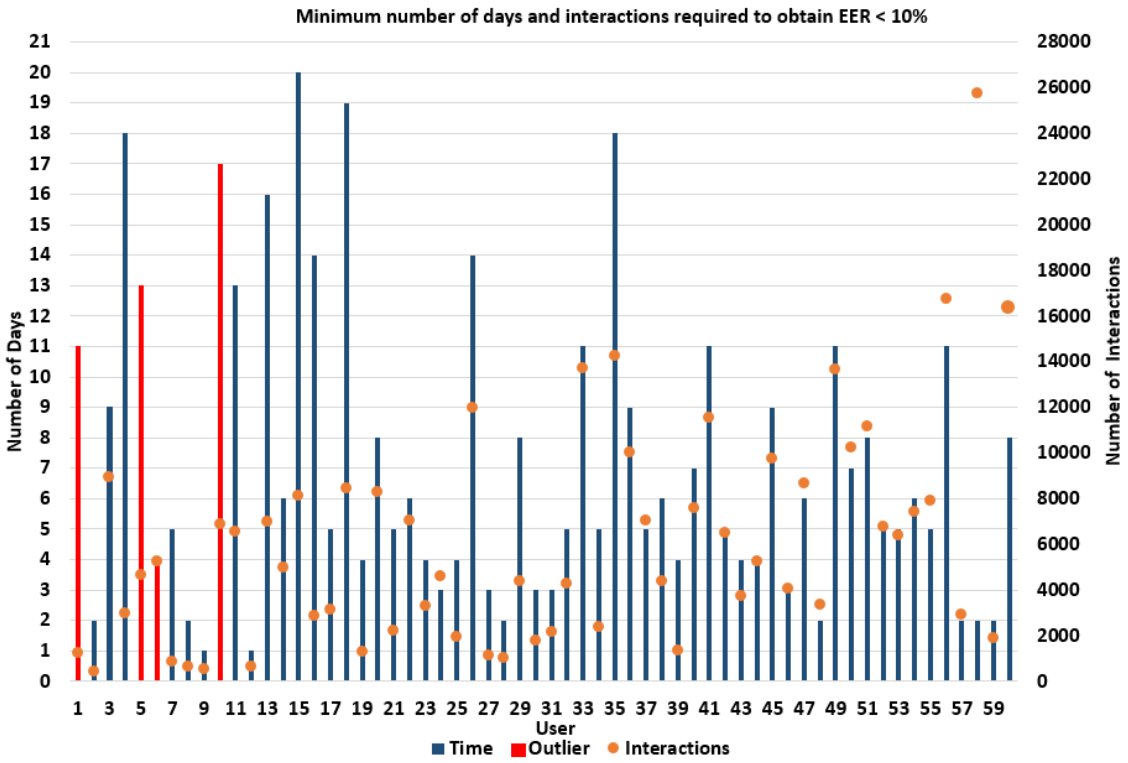


Figure 6.12: Time and volume of data required for generating users' template

Figure 6.12 shows the time and volume of interactions required to generate a user's template are different among users. For example, some users achieved the selected criteria in equal or fewer than two days training with lower interactions compared with other users (i.e., fewer than 1500 interactions), such as users 2, 7, 9, 12, 28, 48, 57, 58 and 59. In comparison, some users needed more than the half of the given period of training with high interactions (more than 17000 interactions) to achieve the goal, such as users 4, 11, 13, 15, 16, 18, 26 and 36. Moreover, some users (outliers), regardless of having had all available period of training, did not achieve the criteria. However, these users had the lowest volume of interactions among most other users (less than 7,000 interactions) such as users 1, 5, 6 and 10. Therefore, the volume of interactions sometime can help the

classifier to achieve a good performance. However, some other users who had few interactions achieved a better performance than users who have many interactions, such as users 9 and 12. This case is a common issue for behavioural based biometrics as users' behaviour tend to change over time and under different external circumstances, which can impact negatively on the sample collection and classification performance. As less control over users and the environment exists, more care is needed when considering their implementation in a verification system. Therefore, it demonstrates that the time and volume factors worked for approximately 94% of the user population. However, these factors may not be always necessary for determining appropriate discriminatory information for users that can help to generate a suitable temple to support the classifier for achieving the correct decision.

6.4 Discussion

The descriptive statistics show there are some strong features that could be investigated to build a sufficient user behaviour profiles. However, some other investigated features showed that it is difficult to discriminate between users of IaaS.

The experimental results of the classification algorithms reveal that cloud infrastructure service users can be strongly identified via their activities with a high degree of accuracy. Shi et al. (2011), Aupy and Clarke (2005) and Yazji et al. (2014) investigated behavioural profiling within a similar environment (computer desktop). However, small datasets were implemented with limited number of participants (8-18), which do not reflect an accurate performance in practical sense.

Additionally, the study by Aupy and Clarke (2005) suffered from a lack of interactions, which led to repeating the number of the previous user interactions to get the richness required interactions to obtain a reliable statistical classification result. The overall outcome was approximately 7% of EER. While this study applied a large dataset containing more than million samples with 60 participants in comparison to the prior art and the performance was better with the best EER of 0.32%, suggesting the usefulness of the proposed technique.

From an individual classifier performance perspective, the experiment showed that the CART algorithm achieved 0.32% EER and outperforms the RF with random sampling and 80/20 training and testing data splitting. This would allow other factors such as time taken to compute, computational overhead, and memory requirements to be considered as part of the selection. In addition, the overall result accuracy of the large volume of data had a positive impact. Users' performance improved with more frequent activities/interactions across both classification algorithms. Moreover, the performance results with random sample selection also achieved better accuracy than the time series selection. This indicates user behaviour is changeable over time and, therefore, care must be taken to ensure appropriate template renewal procedures are developed regularly to maintain levels of performance.

For the time and data volume required to generate a user template, the experiment revealed that five days can be considered as the average time for generating useable user behaviour profiles, as shown in Figure 6.11. However, the five days as a static threshold is not a definitive criterion for creating a user template.

A number of users needed less than five days while others needed more as illustrated in Figure 6.12. In addition, the large volume of data for training is not always guaranteed to perform with better accuracy than the low volume of activity for all users. Therefore, further statistical analysis was applied by selecting users for representing the best and worst cases. Based on Figure 6.12, User 9 was selected as the best case because the user achieved the criteria in the shortest time (one day) and lowest interactions (383 interactions). User 10 was selected as the worst case because he/she did not achieve the criteria over the selected days (17 days) with a high amount of interactions (more than 6,500 interactions). When reviewing on the pattern of the daily usage for the five highest applications/URLs, it is found that User 9 had a consistent pattern of usage during the given days, as shown in Figure 6.13. This could make the classifier identify the user more easily. While User 10 did not seem to have consistent usage as some selected applications/URLs appear within the first few days and disappear within the remained days and vice versa, as shown in Figure 6.14. These changes in user behaviour can have a negative effect on the classifiers' performance because their activities are so diverse.

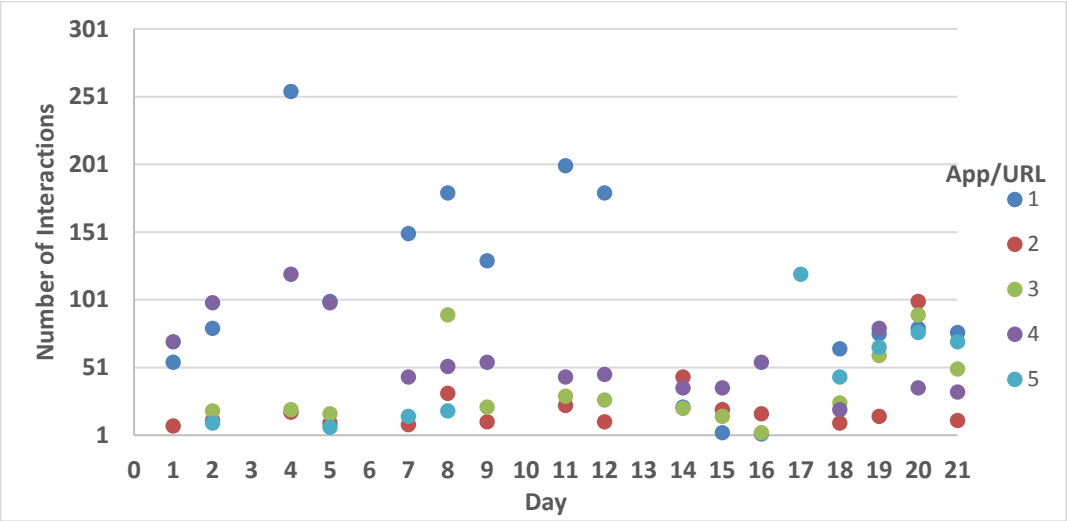


Figure 6. 13: User interaction of User 9

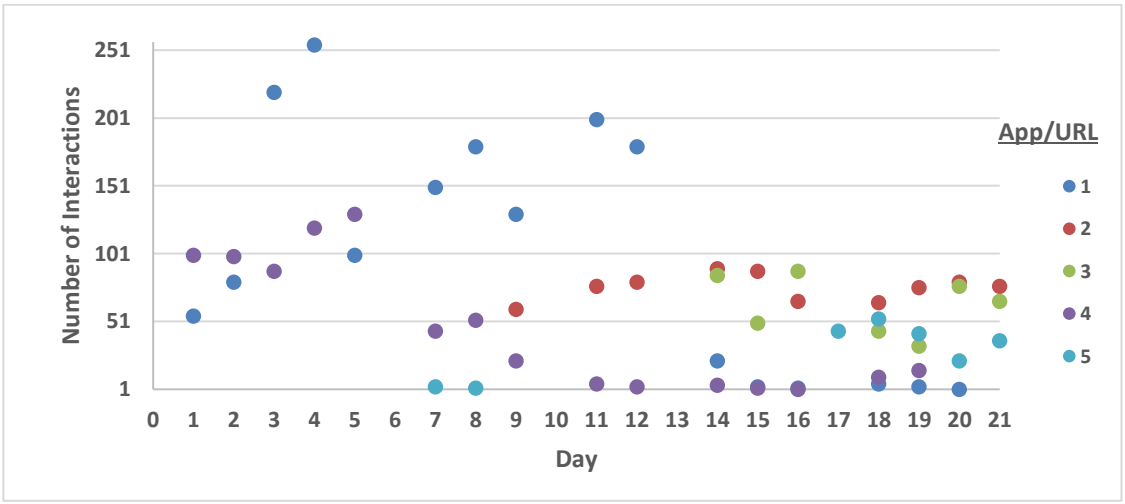


Figure 6. 14: User interactions of User 10

One reason for the worst case could be that part of the dataset was collected from early PhD research students who normally use various applications and websites with no consistency during their initial research period. These variations and changes in user behaviour can negatively affect building an accurate picture of user usage pattern. Therefore, an additional mechanism is required to analyse

the data deeply rather than relying on the time and volume of data alone to provide sufficient discriminatory information for creating these templates.

As this research is aimed at detecting abnormal usage in cloud infrastructure services to identify a misuse, the result has shown that it can be applied the behavioural profiling to achieve this task. However, there are some issues that can be raised by applying this technique such as poor performance when user behaviour is changed. Therefore, this needs to be managed carefully to continue updating the users' templates to remind reflecting on legitimate users. Additionally, this problem can be found in most behavioural biometric techniques and as the aim of this research proposed to use the behavioural profiling as a second factor to continuously verify the identity of the users. Therefore, it can be considered that the behavioural profiling issues are not a major problem for the system because it will not be used to authenticate a user in the point of entry like password or fingerprint (yes/no). It will solely help to identify a misuse in the usage of services after initial login.

6.5 Conclusion

The results successfully demonstrate the ability to correctly distinguish users based on their interactions derived from a simulated cloud infrastructure service environment. Accurate user-behaviour profiles can be built to help distinguish between the normal and abnormal usage. The classification algorithms in the experiment achieved high accuracy, particularly CART. The random sampling achieved significantly better result accuracy than the time series method. In addition, a large number of interactions for training did have a significant impact on

performance. For overall accuracy, participants who achieved high performance were users who had high interactions. Subsequently, the approach proved a highly promising solution for applying user-behavioural profiling as a supporting technique to validate users after the initial point-of-entry authentication. This can contribute and guide the system to identifying a misuse of cloud services in a continuously and friendly manner.

For template generation, the results of further experiments have shown the ability to create sufficient user-behaviour templates. However, the experiments also showed that the time and volume of data are not necessarily key factors for creating these templates. Therefore, there is a need to find criteria to determine for which users the proposed technique can be successfully implemented. Additionally, the template renewal has not been discussed which can be considered as one of the main issues with the behaviour profiling technique that can affect system performance. Therefore, the next chapter will mainly focus on these aspects and suggest some strategies for dealing with them.

7 Discussion

7.1 Introduction

Because of the aforementioned evolution of cloud computing services functionalities along their wide use by users and the increasing rates of cyber attacks, it has become necessary to seek further measures that help secure the data. Transparent user verification has been applied to protect the services and users' data that are hosted in from attacks through tracking and evaluating the user behaviour with these services.

To understand the degree to which applying behavioural profiling can support the verification process to verify the legitimacy of cloud users, practical and operational aspects of SaaS and IaaS studies need to be analysed and compared with other prior art that investigated the same technique to detect illegitimate usage. Moreover, as highlighted in Chapters 5 and 6, user behaviours are changeable over time, which can have a negative impact on the performance of the system in a practical sense. Therefore, this issue needs to be discussed to propose a possible solution that can help the system reduce inconvenient actions such as the wrong alert for legitimate users.

7.2 Comparison with the Prior Art

To understand the performance that was shown by implementing behavioural profiling regarding SaaS and IaaS in Chapters 5 and 6, the proposed approach is compared with the prior work in behavioural profiling. This comparison focuses on the performance has been achieved, the volume and nature of dataset that

being collected and classification algorithms have been employed by these previous studies. This can be helped to understand the feasibility of applying the behavioural profiling technique with cloud services to protect their users account from being compromised. Table 7.1 highlights a copy from analysis of Chapter 4 that illustrates the most related studies that applied user behaviour profiling to protect users from different attack vectors. This include intrusion detection, fraud detection, and authentication across different technologies (e.g., mobile phone system, network, computer system, and web browsing).

Table 7.1: Practical studies of literature review

Author(s)	Activity	Client/Server	#Participants	Performance (%)	Method	Purpose
Moreau et al. (1997)	Telephony	S	#600	DR=90, FRR=10	Supervised Neural Networks	FD
Samfat and Molva (1997)	Mobility	S	#400	DR=82.5, FRR=40	Distance	FD
	Telephony			DR=80, FRR=30	Rule-base	
Hilas et al. (2014)	Telephony	S	#5000	Highest DR=80	Genetic Programing method	FD
Ogwueleka (2009)	Telephony	S	#180	FAR=3	self-Organizing Map and Probabilistic	FD
Qayyum et al. (2010)	Telephony	S	#300	DR=70	Neural Network	FD
Yazji et al. (2011)	Mobility	S	#100	DR=81	cumulative probability and Marko properties of trajectories	IDS
Yazji et al. (2014)	Mobility	S	1-#100	DR=94	cumulative probability and Marko properties of trajectories	IDS
Subudhi and Panigrahi (2015)	Telephony	S	#94	DR=95,TPR=78, FPR=8	SVM	FD
Shi et al. (2011)	Telephony, SMS, Browsing, Mobility	C	#50	DR=95	Probability	Au
Damopoulos et al. (2011)	Telephony, SMS, Browsing	C	#35	DR=98.5, TPR=99.3 FRR=0.7	Bayesian network , RBF, KNN, Random Forest	Au
Li et al. (2014)	Application Activities	C	#76	EER=9.8	Rule base	Au
Fridman et al. (2015)	Text, App, Web and location	C	#200	EER=3	SVM	Au
Aupy and Clarke (2005)	Way of interaction with PC	C	#21	EER= 7	Neural Network (FF-MLP)	Au
Yazji et al. (2009)	File access activity and network events	C	#8	DR=90, FAR=13.7 FRR=11%	K-Means Clustering	Au
Salem and Stolfo (2011)	File access activity	C	#18	FPR=1.1	SVM	Au
Abramson and Aha (2013)	Web Browsing	S	#10	FAR/FRR=24	SVM	Au

*S=Server, C=Client, DR=Detection Rate, EER=Equal Error Rate, FAR=False Accept Rate, FRR=False Reject Rate, TPR=True Positive Rate, FPR=False Positive Rate, FD=Fraud Detection, Au=Authentication, I=Identification

From the above table and to the author's best knowledge, there is a limited number of research on behavioural profiling in cloud computing services to detect misuse. These studies suffered from a lack of real datasets to examine their

thoughts. Doelitzscher et al.'s (2012) study is an example, as they tried to apply behavioural profiling to detect an anomaly in the cloud infrastructure system. The study has several problems; one of the main issues is that they generated simulated data by using software to examine the validity of their approach, which might not represent a real users' interactions in reality. More importantly, the authors focused mostly on the operating system's actions rather than users' actions because they looked at what happened outside the VMs. In contrast, in our study regarding IaaS, a real dataset was gathered from user's interactions with their computer desktops, doing their normal daily activities, which can be considered as a simulated environment of a cloud infrastructure service. Therefore, the collected dataset can act as to the similar users' actions inside the VMs. Therefore, our study was closer to the reality than Doelitzscher et al.'s (2012) study.

Because of the lack of literature in applying behavioural profiling to protect users of cloud computing services from being misused, all studies that illustrated in Table 7.1 are related to creating user behaviour profiles to detect illegal use. This means that the nature of the dataset and the problem would be similar to the dataset and problem of this research. Therefore, as mentioned earlier, the comparison will focus on the dataset that has been used, the system performance that been achieved and the type/nature of classier that been applied in these studies.

From dataset and performance perspective, this research collected a large real dataset with respect to SaaS and IaaS. The SaaS study collected a dataset from 30 users over a six-month period and gained a high performance of average EER equal to 5.8%. Regarding the IaaS study, the dataset was gathered from 60 users

within a three-week period with an EER 0.32% as an average performance. The performance of this research, particularly the IaaS study, achieved the highest accuracy in comparison to all studies illustrated in Table 7.1. However, the table also shows that a number of studies have a larger number of participants than this research and good performance, such as Moreau et al. (1997), Ogwueleka (2009), Hilas et al. (2014), and Fridman et al. (2015). The first three studies achieved an accuracy of a DR ranging from 80% to 90% which can be considered as a high performance based on a large dataset that collected from 600, 300, 5,000 and 200 users respectively to create user behaviour profiles. More importantly, these studies collected their dataset from telephony networks which can gather wide rich information about the users in which help to discriminate among users more accurately. For example, Moreau et al.'s (1997) study generated user behaviour profiles based on toll tickets to extract some important information about calling activities.

Fridman et al.'s (2015) study focused on mobile phones to create user behaviour templates for 200 participants with an EER of 3% as an average performance. They gathered various activities such as text, application, web browsing, and GPS location activity. These activities can give wide information about the users, which can help distinguish the usage of each user easily.

In comparison with this research, only three features were included: time, file type and event for SaaS study. Also, these features are generated from users' interactions during interacting with a single application only (Dropbox activity). Regarding the IaaS study, three features have also been used which are: time,

application and web browsing. This limited feature vector affects the richness of users' information which would have a negative impact on the discrimination among users. However, with all these restrictions, the research achieved a high performance particularly IaaS study. Moreover, increasing the dimensionality of the feature vector will lead to an increase in the computation of classification algorithms intensively. In reality, this may have a negative effect on the speed decision for detection the imposter.

More importantly, the first three studies calculated only the FRR metric, which was high (negative). However, the studies did not show the FAR, which is not fair because both these metrics rely on a selected threshold. Therefore, as the value of the determined threshold increases, the value of FAR decreases and vice versa. Thus, without computing both metrics, it does not reflect a real accuracy of a biometric system.

There are some similar studies to IaaS including the environment of collecting a dataset and the research problem. These studies are Aupy and Clarke (2005), Yazji et al. (2009), and Salem and Stolfo (2011). The studies focused on generating user behaviour profiles from desktop computers used to detect any illegal access to the device. However, one of the main limitations of these studies is the datasets used contain a limited number of participants ranging from 8 to 18 users. On the other hand, the IaaS study included 60 users. A small dataset would not reflect an accurate performance in a practical sense. Although these studies collected the small dataset, the best performance was 7% of EER which is higher than the performance of the IaaS study (0.32% of EER).

Therefore, based on the above analysis, it can be considered that this research is in line with the highest results that are achieved in the related works.

From the classification algorithms perspective, Table 7.1 shows that most of the studies were conducted by implementing a single classifier to examine the validity of applying behaviour profiling technique to detect several types of attacks with different technologies. For example, a neural network algorithm and SVM are the most frequent classification algorithms that have been used. However, this research implemented four classifiers (SVM, RF, CART, and FF MLP neural network) to examine the effect of these different algorithms and to determine the best one. The study showed that the CART was the best classifier based on the accuracy and speed of computation compared with the three selected classifiers, as shown in Table 7.2 (copied from Table 5.7 in Chapter 5).

Table 7.2: Performance of classification algorithms

Classifier	Time (D:H:M:S)	EER (%)
SVM	00:04:33:08	20.27
RF-25 trees	00:00:50:15	9.93
FF MLP Neural Network-65 neurons	02:02:40:55	6.98
CART	00:00:10:25	6.02

This research also shows that the FF MLP Neural Network and SVM have a number of drawbacks. Firstly, The FF MLP neural network is a more complex classifier compared with the three classifiers that are used in this research. It is changing and time-consuming to find the best parameters' combinations that make the model converge. For example, the number of neurons needs to be examined with the number of iterations to select the best number of neurons that

can achieve the best performance. As a result, it requires more computational power and time to achieve the task. As shown in Table 7.2, it took more than two days to train and test the given data except for the setting time, knowing the dataset was small (SaaS dataset) comparing with a large dataset of IaaS study. Therefore, with a large dataset and feature factor, it might take months to tune those parameters and to train and test the users' interactions. Whereas, the CART is not required for any setup, with default setting can achieve better performance in a short time (Wu et al. 2008). Regarding SVM, it achieved the lowest accuracy with this research, and it did not work with the dataset of IaaS because it can work with a limited volume of data. Therefore, this type of the classifier might be difficult to be implemented in a practical sense with a large number of scale problems. For example, a dataset of cloud computing services can be a large number of users with a vast number of interactions.

7.3 Enrolment and Template Renewal

As highlighted in chapters 5 and 6, a number of issues need to be managed which can face the proposed approach in a practical sense and affect the system performance. One of the most important issues is user behaviour changes over time because it can have a high impact on the system performance. This was clear with a number of users who performed particularly poorly when applying time series rather than bootstrap for the data selection process, as shown in SaaS and IaaS study (in section 5.3.2 and 6.3.3 respectively). A thorough analysis of some users' features showed that the changes in user behaviour over time can have a negative impact on the system performance. For example, the lowest performances—User 8 from the dataset of SaaS study compared to all other users—

could be caused by the drastic changing in his/her behaviour. Figure 7.1 shows some of these changes during the selected days for event activity.

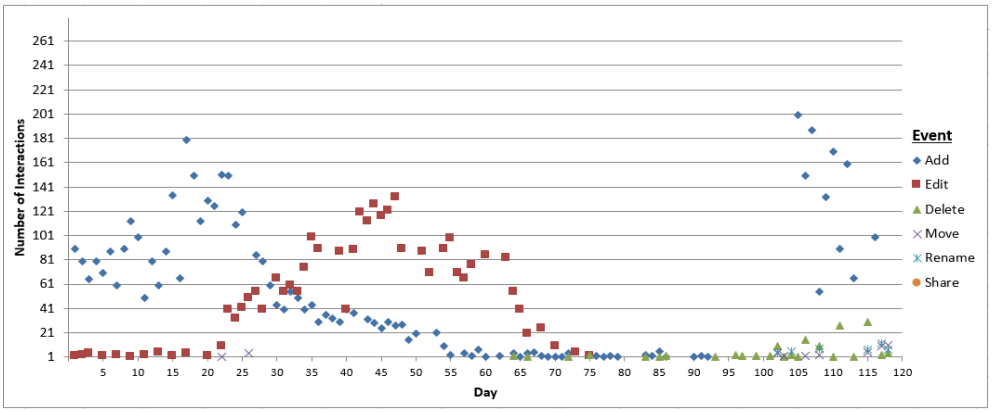


Figure 7.1: User interactions of User 8

Moreover, user 5 and 10 in the IaaS study did not reach the selected criteria (10% of EER) across all given days, as shown in Figure 6.12 in section 6.3.3. When looking more deeply on the five selected applications/URLs, which have the highest interactions over the given period compared with other applications/URLs, it is shown that the users' behaviours with these applications/URLs are changed over the selected period, as illustrated in Figures 7.2 and 7.3.

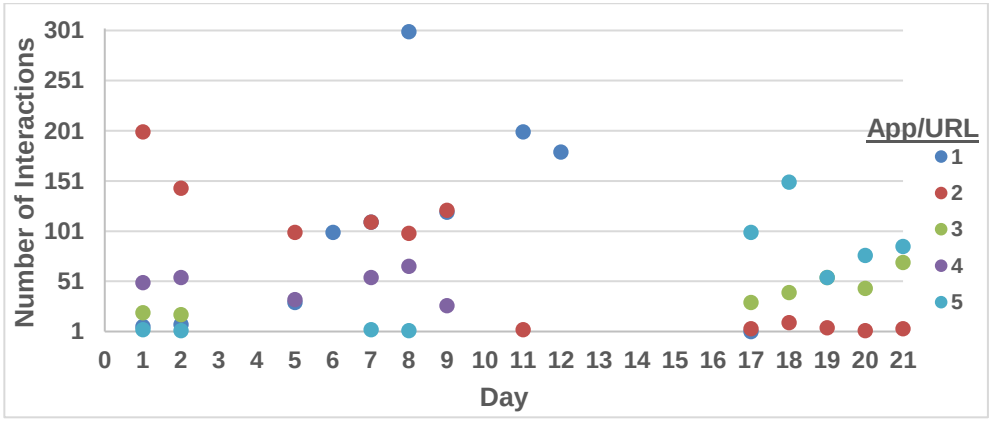


Figure 7.2: User interactions of user 5

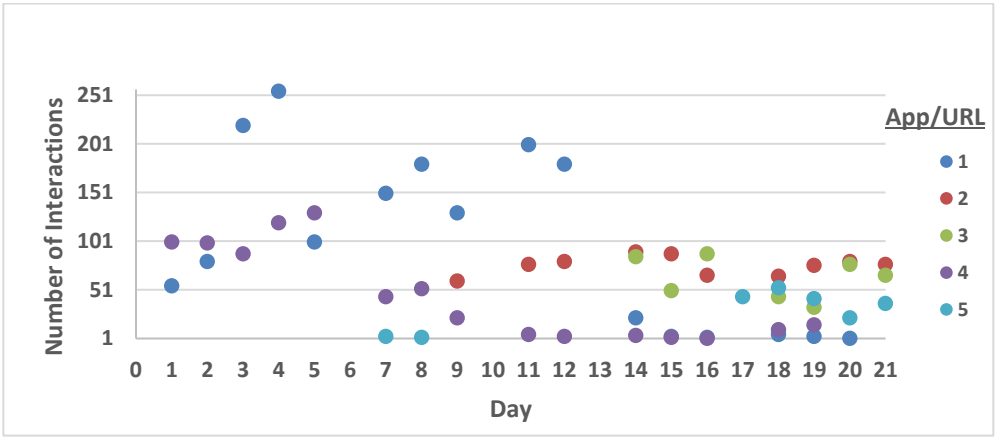


Figure 7.3: User interactions of user 10

From the all above figures, it can be shown that some user's activities were changeable over time and some of them appeared with few interactions then increased/disappeared such as 'edit' event in Figure 7.1 and the application/URL number 1 in Figure 7.2. While other activities did not appear in the first period, they increased after a while later such as 'delete' event in Figure 7.1 and the application/URL number 3 in Figure 7.3. Therefore, if a classifier is trained on the specific user patterns, any change in these patterns over time can make the classifier struggle to recognise these changes during the verification process. The

system could consider these changes in user behaviour as illegitimate usage because it would affect the system's performance. As the proposed technique serves to monitor misuse, this might stop the current legitimate user who uses the service or at least these changes could make the usage of the service inconvenient because of the increasing number of alarms. Therefore, an adaptive renewal of user's template is needed to include the most recent changes in users' behaviour as well as the old users' patterns that do not become related to the current user behaviour needs to be removed from the template.

Therefore, a procedure for renewing user's template was implemented to adopt new behaviour change. This is described as follows: firstly, a period of time can be set up for the training stage which should precisely reflect the current user to generate the initial template. Then, this template can be updated in a timely fashion which will be selected as a daily basis in this procedure. Therefore, after generating the initial template and to keep including the most recent user's behaviour, the day after will be used for testing and then updating the template. Additionally, the first day of usage will be removed from the user's template. The same procedure will be repeated on the next day and so on. To do so, the dataset of IaaS study was selected as it had more user interactions than the SaaS study. Additionally, based on the experimental results that were achieved in IaaS, Figure 6.11 (section 6.3.3) shows that five days can be selected as an initial users' template, as the accuracy of most users after these days remained relatively stable. Therefore, as a dynamic template renewal procedure, the sixth day's data will be used for testing and updating the template while the first day of usage will be removed from the template. Then the seventh day of usage will be used for testing

and updating and the second day of the user's usage will be removed and so on. Figure 7.4 shows the process of the suggested procedure.

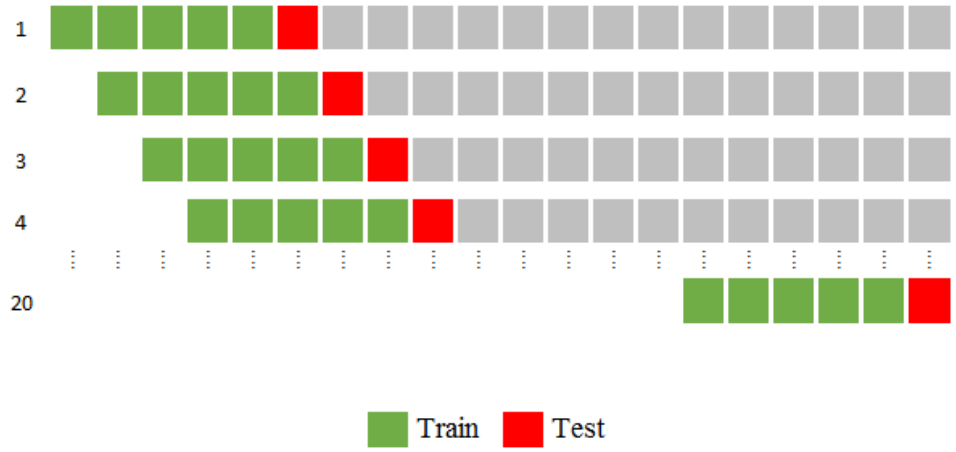


Figure 7.4: Dynamic template renewal procedure

The CART classifier was chosen to examine the idea of this procedure, as it achieved the best outcome in SaaS and IaaS studies. Figure 7.5 illustrates the overall performance of all users across all selected days.

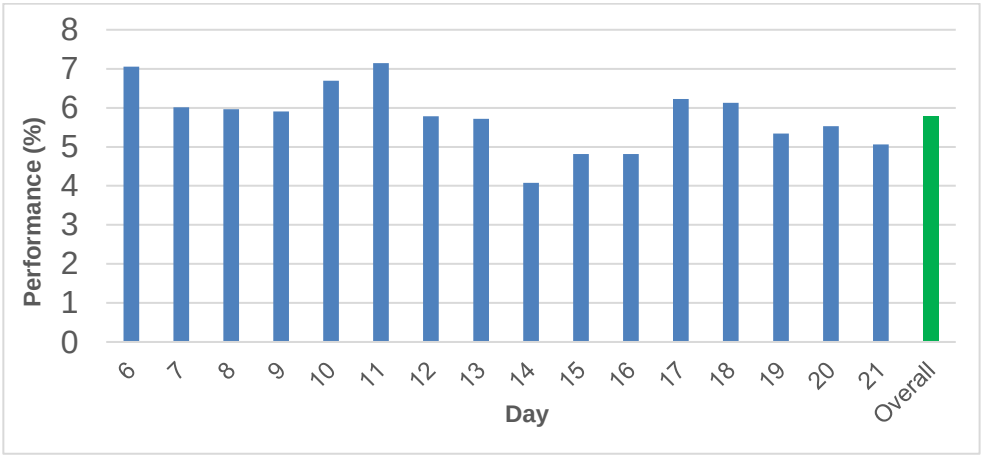


Figure 7.5: Average of users' performance of template renewal procedure

The overall performance in Figure 7.5 is encouraging in favour of the idea of using the dynamic renewal of users' templates. Comparing with the overall experimental results of IaaS study that used static users' templates, as shown in Table 7.3, the overall performance of the dynamic template renewal achieved a better result with an overall EER of 5.77%.

Table 7.3: Overall users' performance with static templates

EER (%) based on the volume of data			
Classifier	Time-series selection		
	50/50	66/34	80/20
CART	8.51	7.35	6.55

*Note: this table is extracted from Table 6.3 in section 6.3.2

Table 7.3 shows that although different sizes of static users' templates were used, the template renewal achieved higher performance than all these static temples. This means that template renewal, which includes the most recent user's behaviour can have a positive effect on the performance of the system more than increasing the size of templates.

However, Figure 7.5 shows that the overall daily basis accuracy results of the template renewal slightly fluctuates across the most chosen period. This can be considered as normal because sometimes user behaviour might change sharply. As the current template does not include this change yet, this can affect negatively on the performance of the system. Figure 7.6 shows the performance of each user individually across the selected days.

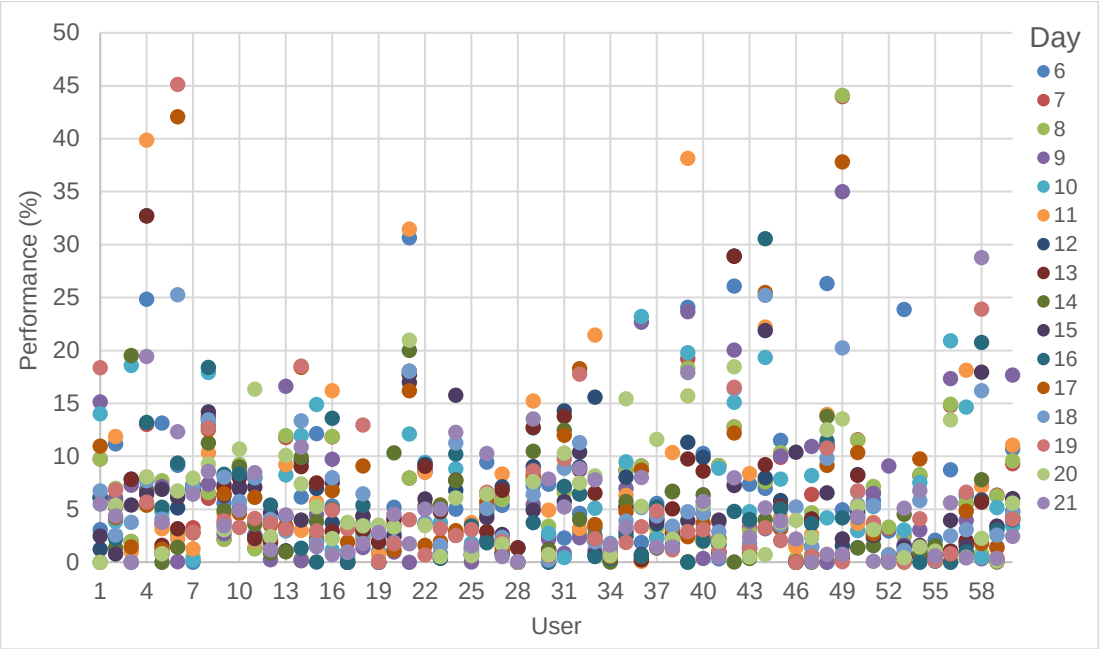


Figure 7.6: Users' performance of dynamic template renewal

As shown in Figure 7.6, most users' performance based on daily usage have a good performance, which are less than 10% of EER. However, some users had a low performance in some days such as User 6 in days 17 and 19 and User 49 in days 7 and 8 with an EER of 42.09%, 45.15%, 43.99% and 44.13%, respectively. As mentioned, the unexpected change in user behaviour could cause this problem. Therefore, a short period could be used, such hourly, for updating the template to manage such issue.

However, this users' template renewal procedure may include impostor's data over the time. As seen in Figure 7.6, the low performance might not mean an unexpected change in the user's behaviour, it might be impostor's data. In this case, the new template might include impostor's interactions over the time and the system will deal with these interactions as legitimate interactions in the future.

This is an uneasy task to be solved, as dealing with user's template renewal using single behavioural biometrics is more challenging than dealing with a multi-biometric system. Nevertheless, a majority voting system can be employed to alleviate this issue by relying on a group of user interactios to make a decision for updating the template or not; if most of the interactions perform well, all these interactions can be considered as legitimate interactions and hence be included in the template renewal process; otherwise, the system would reject these interactions.

However, the majority voting technique might not be also a perfect solution for solving the issue of high degree of change in user's behaviour as in such situation most interactions would be dissimilar with the user behavioural templates. This means that most of the new interactions could not be detected with high confidence to the class to which they belong. Therefore, these interactions would not be included in the template renewal, whereas all interactions should be considered as legitimate interactions. Additionally, the approach cannot be also employed with users who have low interactions during a long period of time, as seen with users of Dropbox in section 5.3.1.3. Some users did not use Dropbox services frequently. Therefore, it is difficult to apply this approach with these types of users even though the results of majority voting of these interactions are positive because the time between one interaction and another is too long which leading it to probably be impostor interactions. Therefore, an intelligent system is needed to deal with these situations carefully to avoid including illegitimate interactions whilst renewing users' templates.

7.4 Conclusion

The intense analysis and comparison between this research and the previous studies, that have applied the behavioural profiling technique, showed that this research is in line with the highest results achieved in the related works. The selected approach proved a highly promising solution for applying user-behavioural profiling as a supporting technique to verify users after the initial login. This can help the system identify misuse of cloud users' usage in a continuous and friendly manner.

Regarding renewal of users' templates, the results of further experiments have shown that dynamic template renewal regularly can achieve a better performance of using static templates, as it will include the most recent user behaviour. However, there is a high risk of including impostor data within legitimate templates. Therefore, an intelligent system is needed to make more measurements that can help to reduce the risk of impostor's interactions being included with the refreshing users' template.

8 Conclusions and future work

This chapter concludes the main achievements of the research and discusses the research's limitations and obstacles. It also highlights the potential areas for further studies within the security field of cloud computing services.

8.1 Achievements of the research

Overall, all the aims of the research of improving the security of cloud computing services initially set out in Chapter 1 have been achieved, through applying a series of practical experiments on two novel datasets. The key achievements of this research are:

1- **Design a series of experiments to explore the feasibility of deploying behavioural-based profiling on the top layer of cloud computing services (SaaS).**

This goal achieved by developing a novel series of experimental studies on user application activities within cloud computing services (cloud storage services). The experiments focused on evaluating user behaviour profiles within the top application level of cloud services (SaaS). A novel dataset of users' interactions with a cloud storage service (Dropbox) was collected from 30 participants. By applying the descriptive statistical approach, several features were obtained and analysed to show the potential success of building user behaviour profiles. Then, these profiles were employed and evaluated by more complex solutions using four classification algorithms (Neural Networks, SVM, RF and CART). Various as-

pects were examined including the impact of the data volume, type sample selection, and the nature of the classifier on the system's performance. The biggest volume of data with random sample selection showed a positive impact on the accuracy of the selected classifier, as it can cover the patterns of most users. Importantly, the most optimal classifier of the experimental study was one of the decision tree approach (CART) that can be used to build successful user behaviour profiles within the cloud computing environment. The performance of this study is encouraging and shows the ability to identify misuse within cloud storage services via the behavioural profiling technique.

2- **Design a series of experiments for investigating the feasibility of deploying behaviour-based profiling on underlying layers of cloud computing services (IaaS).**

The second main contribution was applying behavioural profiling on an underlying infrastructure cloud model (IaaS) with a larger dataset than the dataset of the previous study. As there were no public datasets available for IaaS because of privacy and security concerns of cloud providers, software was developed to capture real users' interactions from computer desktops for a significant period of time without any conditional control on the users. This resulted in having a private real dataset that was collected from 60 users by installing the software on the participants' computers. The obtained dataset is considered as an image for users' interactions with IaaS. The volume of the dataset of this study was more than ten times larger than the dataset of the previous study. Although most of the

methodologies of this study were similar to those of previous studies, the performance of this study was better than that of previous studies in all aspects; overall; it is higher than that obtained by earlier studies. In addition, this study's performance was better than prior work within a similar environment (computer desktop) although the dataset was bigger. Finally, further investigation was made in this study, which focused on the volume of data needed for each user to build a suitable behaviour profile. This investigation showed that the discriminatory information of these profiles can be affected by the time and the volume of data.

3- Propose a continuous verification approach that can keep updating users' template dynamically in order to mitigate the effect of user behaviour change over time on the performance of the system.

This objective was achieved by examining the effect of user behaviour change over time on the performance and apply a periodical model adoption approach to keep users' templates updated dynamically to continue reflecting legitimate users. The result of this approach got better performance than the performance of the previous experiment of the static user's templates.

8.2 Limitations of research

Whilst the objectives of this research programme have been achieved, a number of limitations associated with the research can be identified. They are summarised as follows.

1- From an experimental dataset perspective, there are a number of limitations. Firstly, the number of participants in the SaaS and IaaS studies were limited (30 and 60 participants respectively), as mentioned in (Chapters 5 and 6). These datasets also had few activities when compared with further activities recorded by cloud providers, which have other activities and their metadata such as IP address, type of browser ,and operation system or device. Additionally, the dataset of the second study (Chapter 6) does not reflect a complete image of the real-world scenarios of users' interactions with cloud infrastructure services. In addition, the duration of this dataset was limited. More participants and activities with a longer profile period would better provide a more accurate measure of accuracy. This can be allowed to understand the ability of the system to detect misuse with a large number of users and to understand the change in user behaviour over time more deeply. This can be helped to set up a possible solution to manage any unexpected cases that can face the system in the future.

2- There is a concern about the performance of the available dataset (particularly Dropbox dataset) when moving forward in practical use (real world) based scenario in terms of speed detection with the volume of data that are needed to get a reliable outcome. This might make the proposed technique inconvenient for users, such as triggering the wrong alert to legitimate users.

3- Because of the large number of records (more than one million of application/websites (observations) collected in the IaaS study, a huge amount of

computing resources are needed to fully examine all the experimental settings. For instance, the study excluded a neural network algorithm, as it required more than one month finishing the training task. Therefore, selected machine learning algorithms (RF and CART) were included in the second study that fit with the available tested resource.

4- Applying a short time window to verify the users continuously might be difficult because users who do not have enough interactions within a short time might not achieve a good accuracy. However, given a long time window of verification might enable an impostor to misuse the service within 1 or 2 minutes without necessarily affecting the verification result; this can cause two serious problems. Firstly, the service can be abused; secondly, all impostor samples will have access as legitimate user samples. As a result, over time, the users' profile will have impostor samples in their templates. Therefore, the proposed idea of renewing templates could not be applied to users having limited interactions, especially if the time between samples was too long. In this case, the idea of the majority samples (legitimate/illegitimate samples) could not be implemented.
5- Limited number of machine leering algorithms was investigated which might be applying other algorithms could give a better performance.

8.3 Scope for future work

The main aim of this research was to improve the security of cloud computing services through continuous verification of the validity of the current user. For further research and exploration, several scenarios can be investigated as follows:

1- The number of participants and duration of data collection can be expanded to collect a large dataset. This would enable better understanding of the model's changes in the real world, such as user behavioural changes.
2- Further experimental investigations are required to focus on many aspects. For instance, developing mechanisms to understand the nature of the user activities more deeply to make sure appropriate user-behaviour profiles can be generated including adaptive dynamic feature selection and when and how template renewal should be undertaken.
3- Developing a behaviour profiling framework prototype on real cloud computing services that have the ability to verify users during interaction with their services including examining of real impostor dataset could be implemented. This would allow the researchers to evaluate the technique in a practical sense from various aspects such as decision accuracy, response time and processing requirement. Also, more accurate participant feedback can be collected.
4- To ensure the users' behaviour change over time is included in their templates with no impostor samples, an intelligent approach needs to be developed in particular for the users with low interactions. For example, Microsoft and Google are examples of having multi-levels of cloud services (SaaS, PaaS, and IaaS), meaning a user can have different services within a cloud provider. To elaborate, the user can have IaaS with Microsoft as a provider and the same user can use SaaS with the similar provider. This can help to build a stronger behaviour profile to the user. Therefore, an

intelligent verification system can be developed to build multi-instance behaviour profiling based on different levels of services.

5- An intelligent system should be developed to determine the users who can use the suggested technique successfully because the system might be inconvenient to some users who have limited access to the services because limited samples might not acquire a reliable result.

8.4 The future of behavioural profiling for verification users of cloud services

Cloud computing is going to continue to offer various services to individuals and organisations. Customers rely on these services to complete personal and business jobs daily. They can build and run projects, browse and buy products, send and receive emails, store confidential information, transfer money and communicate with friends. However, these cloud services are becoming exposed by cybercriminals even though security controls were in place and dedicated security teams were allocated. At this stage, although authentication is a necessity, it can be argued that it is not longer a strong option to fully protect users from attacks such as misuse.

This research highlights the essential need for a new robust and reliable security mechanism to verify a user's identity and detect misuse actions to individuals during the usage of the cloud services. This a new approach of user verification is dedicated to cloud service providers that can offer a centralised transparent and continuous verification mechanism of user legitimacy who interact with their services.

In the near future, continuous user identity verification for cloud computing services after initial login will become a vital aspect to protect their customers from misuse. Providers of these services will have to apply multi-security techniques to offer a strong protection for their users through a continuous and transparent verification to provide both the security and usability.

References

Abramson, M. and Aha, D.W., 2013, May. User Authentication from Web Browsing Behavior. In FLAIRS conference (pp. 268-273).

Adair, K.L., Parthasaradhi, S.T. and Kennedy, J., 2008. Real world evaluation: Avoiding pitfalls of fingerprint system deployments. *Lumidigm, Whitepaper*.

Agrawal, B., Wiktorski, T. and Rong, C., 2017. Adaptive real-time anomaly detection in cloud infrastructures. *Concurrency and Computation: Practice and Experience*, *29*(24), p.e4193.

Aleem, A. and Ryan Sprott, C., 2012. Let me in the cloud: analysis of the benefit and risk assessment of cloud platform. *Journal of Financial Crime*, *20*(1), pp.6-24.

Alguliev, R. and Abdullaeva, F., 2014. Illegal Access Detection in the Cloud Computing Environment. *Journal of Information Security*, *5*(02), p.65.

Alhanahnah, M., Bertok, P., Tari, Z. and Alouneh, S., 2018. Context-aware multifaceted trust framework for evaluating trustworthiness of cloud providers. *Future Generation Computer Systems*, *79*, pp.488-499.

Annappaian, D.H. and Agrawal, V.K., 2015. Cloud services usage profile based intruder detection and prevention system: intrusion meter. *Transactions on Networks and Communications*, *2*(6), pp.12-24.

Aupy A and Clarke N.,2005. User Authentication by Service Utilisation Profiling. *Advances in Network and Communications Engineering 2*. School of Computing, Communications & Electronics, University of Plymouth, 18.

AWS, 2018. *Amazon Web Services (AWS) - Cloud Computing Services*. Available at: https://aws.amazon.com/ (accessed 15/05/18).

Becker, R.A., Volinsky, C. and Wilks, A.R., 2010. Fraud detection in telecommunications: History and lessons learned. *Technometrics*, *52*(1), pp.20-33.

Bennett A., 2017. *8 Public Cloud Security Threats To Enterprises In 2017*. Available at: https://www.comparethecloud.net/articles/8-public-cloud-security-threats-to-enterprises-in-2017/ (accessed 14/05/18).

breach level index, 2017. *Data Breach Statistics by Year, Industry, More - Breach Level Index*. *Breach Level Index*. Available at: http://breachlevelindex.com/.(accessed 13/05/18).

Burge, P. and Shawe-Taylor, J., 1997. Detecting cellular fraud using adaptive prototypes. *Proc. AI Approaches to Fraud Detection and Risk Management*, pp.9-13.

Buschkes, R., Kesdogan, D. and Reichl, P., 1998, December. How to increase security in mobile networks by anomaly detection. In *Computer Security*

Applications Conference, 1998. Proceedings. 14th Annual (pp. 3-12). IEEE.

Buyya, R., Vecchiola, C. and Selvi, S.T., 2013. *Mastering cloud computing: foundations and applications programming*. Newnes.

Cameron D., 2014. *Apple knew of iCloud security hole 6 months before Celebgate*. *The Daily Dot*. Available at: https://www.dailydot.com/debug/apple-icloud-brute-force-attack-march/ (accessed 27/02/18).

Ceccarelli, A., Montecchi, L., Brancati, F., Lollini, P., Marguglio, A. and Bondavalli, A., 2015. Continuous and transparent user identity verification for secure internet services. *IEEE transactions on dependable and secure computing*, *12*(3), pp.270-283.

Centres SD, 2017. *Cloud platform performance - mid-term report _ Space Data Centres*. . Available at: https://www.spacedatacentres.co.uk/cloud-platform-performance/ (accessed 27/02/18).

Stormann C., 1997. Fraud management tool : evaluation report. 1–30.

Chen, D. and Zhao, H., 2012, March. Data security and privacy protection issues in cloud computing. In *Computer Science and Electronics Engineering (ICCSEE), 2012 International Conference on* (Vol. 1, pp. 647-651). IEEE.

Chen P, Desmet L, Huygens C, Chen P, Desmet L, Huygens C, Study A, Persistent A and Bart T., 2016. A Study on Advanced Persistent Threats. *Springer, Berlin, Heidelberg* 8735: 63–72.

Shaikh, F.B. and Haider, S., 2011, December. Security threats in cloud computing. In *Internet technology and secured transactions (ICITST), 2011 international conference for* (pp. 214-219). IEEE.

Chu, C., Hsu, A.L., Chou, K.H., Bandettini, P., Lin, C. and Alzheimer's Disease Neuroimaging Initiative, 2012. Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *Neuroimage*, *60*(1), pp.59-70.

Cisco, 2018. *Cisco Global Cloud Index : Forecast and Methodology , 2016-2021*. *White Paper*. Available at: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns1175/Cloud_Index_White_Paper.html#wp9000816.

Clarke, N.L., Furnell, S.M. and Reynolds, P.L., 2002, November. Biometric authentication for mobile devices. In *Proceeding of the 3rd Australian information warfare and security conference* (pp. 61-69).

Clarke, N., 2011. *Transparent user authentication: biometrics, RFID and behavioural profiling*. Springer Science & Business Media.

Cloud Security Alliance, 2016. The Treacherous 12 Cloud Computing Top Threats in 2016. *Security* (February): 1–34. Available at:

http://www.cloudsecurityalliance.org/topthreats.%5Cnhttp://www.cloudsecurityalliance.org. (accessed 20/02/17)

CloudRAIL, 2017. *Cloud Storage Report 2017 - Dropbox Loses Market Share But is Still the Biggest Provider on Mobile - CloudRail*. . Available at: https://blog.cloudrail.com/cloud-storage-report-2017/ (accessed 19/11/17).

Crawford, H. and Renaud, K., 2014. Understanding user perceptions of transparent authentication on a mobile device. *Journal of Trust Management*, *1*(1), p.7.

Crowd, 2018. *Cloud Security Report 2018*. . Available at: https://www.alertlogic.com/press-releases/alert-logic-releases-2015-cloud-security-report/ (accessed 14/05/18).

Damopoulos, D., 2013. Anomaly-Based Intrusion Detection and Prevention Systems for Mobile Devices: Design and Development. *University of the Aegean*.

Danny P., 2013. *Green light for National Grid's cloud move*. *Computing.Co.UK*. Available at: http://www.computing.co.uk/ctg/analysis/2257295/green-light-for-national-grid-s-cloud-move (accessed 15/05/18).

Yadron D., 2016. *Hacker collects 272m email addresses and passwords, some from Gmail | Technology | The Guardian*. *Theguardian*. Available at: https://www.theguardian.com/technology/2016/may/04/gmail-yahoo-email-password-hack-hold-security (accessed 10/03/18).

Damopoulos, D., Menesidou, S.A., Kambourakis, G., Papadaki, M., Clarke, N. and Gritzalis, S., 2012. Evaluation of anomaly-based IDS for mobile devices using machine learning classifiers. *Security and Communication Networks*, *5*(1), pp.3-14.

Bijwe, S. and Ramteke, P., 2015. Database in Cloud Computing-Database-as-a Service (DBaas) with its Challenges'. *International Journal of Computer Science and Mobile Computing*, *4*(2), pp.73-79.

Doelitzscher F, Reich C, Knahl M, Passfall A and Clarke N., 2012. An agent based business aware incident detection system for cloud environments. *Journal of Cloud Computing: Advances, Systems and Applications* 1(1): 9.

Dropbox, 2017. *About - Dropbox*. *Dropbox*. Available at: https://www.dropbox.com/about (accessed 07/11/17).

ebay inc, 2010. *BREAKING: eBay and Microsoft Announce Cloud Computing Agreement #WPC10 - eBay InceBay Inc*. . Available at: http://blog.ebay.com/breaking-ebay-and-microsoft-announce-cloud-computing-agreement-wpc10/ (accessed 27/04/15).

Griffith, E., 2014. Who's winning the consumer cloud storage wars. Retrieved from. Available at: http://fortune.com/2014/11/06/dropbox-google-drive-microsoft-onedrive/ (accessed 20/11/17).

Fernandes, D.A., Soares, L.F., Gomes, J.V., Freire, M.M. and Inácio, P.R., 2014. Security issues in cloud environments: a survey. *International Journal of Information Security*, *13*(2), pp.113-170.

Florentine S., 2016. *Cloud adoption soars, but integration challenges remain | CIO*. *Cio*. Available at: http://www.cio.com/article/3018156/cloud-computing/cloud-adoption-soars-but-integration-challenges-remain.html (accessed 18/06/18).

Forbes, 2015. *Roundup Of Cloud Computing Forecasts And Market Estimates, 2015*. . Available at: http://www.forbes.com/sites/louiscolumbus/2015/01/24/roundup-of-cloud-computing-forecasts-and-market-estimates-2015/. (accessed 27/04/15).

Fowler, G.A. and Worthen, B., 2009. The internet industry is on a cloud-whatever that may mean. *The Wall Street Journal*, *253*(70), pp.1-10.

Fridman, L., Weber, S., Greenstadt, R. and Kam, M., 2017. Active authentication on mobile devices via stylometry, application usage, web browsing, and GPS location. *IEEE Systems Journal*, *11*(2), pp.513-521.

Gamboa, H. and Fred, A., 2004, August. A behavioral biometric system based on human-computer interaction. In *Biometric Technology for Human Identification* (Vol. 5404, pp. 381-393). International Society for Optics and Photonics.

Hogben, G., 2010. *ENISA Briefing: Behavioural Biometrics*. Technical report, European Network and Information Security Agency (ENISA).

Grosser, H., Britos, P. and García-Martínez, R., 2005, June. Detecting fraud in mobile telephony using neural networks. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems* (pp. 613-615). Springer, Berlin, Heidelberg.

Gupta, B.B., Yamaguchi, S. and Agrawal, D.P., 2018. Advances in security and privacy of multimedia big data in mobile and cloud computing. *Multimedia Tools and Applications*, *77*(7), pp.9203-9208.

Gupta, S., Kumar, P. and Abraham, A., 2013. A profile based network intrusion detection and prevention system for securing cloud environment. *International Journal of Distributed Sensor Networks*, *9*(3), p.364575.

Hall, J., Barbeau, M. and Kranakis, E., 2005, August. Anomaly-based intrusion detection using mobility profiles of public transportation users. In *Wireless And Mobile Computing, Networking And Communications, 2005.(WiMob'2005), IEEE International Conference on* (Vol. 2, pp. 17-24). IEEE.

Harikrishan V., 2015. *Cloud Computing : Levels (IaaS, PaaS, SaaS) and Deployment models (Public, Private, Hybrid) - KnowledgeBlob*. . Available at: http://knowledgeblob.com/technology/cloud-computing-levels-iaas-paas-saas-

and-deployment-models-public-private-hybrid/. (accessed 27/04/15)

Hashem, I.A.T., Yaqoob, I., Anuar, N.B., Mokhtar, S., Gani, A. and Khan, S.U., 2015. The rise of “big data” on cloud computing: Review and open research issues. *Information systems*, *47*, pp.98-115.

Hebah E., 2011. User Profiling in Telecommunication Fraud Detection.

Hemparuva, R.J.C., Simon, S.P., Kinattingal, S. and Padhy, N.P., 2018. Geographic information system and weather based dynamic line rating for generation scheduling. *Engineering Science and Technology, an International Journal*.

Hilas, C.S., Kazarlis, S.A., Rekanos, I.T. and Mastorocostas, P.A., 2014. A genetic programming approach to telecommunications fraud detection and classification. In *Proc. 2014 Int. Conf. Circuits, Syst. Signal Process. Commun. Comput* (pp. 77-83).

Hilas, C.S. and Sahalos, J.N., 2007, September. An application of decision trees for rule extraction towards telecommunications fraud detection. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems* (pp. 1112-1121). Springer, Berlin, Heidelberg.

Hilas, C.S. and Sahalos, J.N., 2005, October. User profiling for fraud detection in telecommunication networks. In *5th International conference on technology and automation* (pp. 382-387).

Hu, Y., Frank, C., Walden, J., Crawford, E. and Kasturiratna, D., 2011, April. Profiling file repository access patterns for identifying data exfiltration activities. In *Computational Intelligence in Cyber Security (CICS), 2011 IEEE Symposium on* (pp. 122-128). IEEE.

Hussain, M. and Abdulsalam, H., 2011, April. SECaaS: security as a service for cloud-based applications. In *Proceedings of the Second Kuwait Conference on e-Services and e-Systems* (p. 8). ACM.

Chandra, J.V., Challa, N. and Hussain, M.A., 2014. Data and information storage security from advanced persistent attack in cloud computing. *International Journal of Applied Engineering Research*, *9*(20), pp.7755-7768.

Jadeja, Y. and Modi, K., 2012, March. Cloud computing-concepts, architecture and challenges. In *Computing, Electronics and Electrical Technologies (ICCEET), 2012 International Conference on* (pp. 877-880). IEEE.

Jain, A.K., Flynn, P. and Ross, A.A., 2007. Handbook of Biometrics: Automatic Forensic Dental Identification.

Jain, A.K., Ross, A. and Prabhakar, S., 2004. An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology*, *14*(1), pp.4-20.

Jain, P. and Pandey, U.K., GPS Based Authentication Mechanism in Cloud Computing.

Jake, 2018. IaaS, PaaS and SaaS: The Cloud Comparison Guide [2018 Update]. *RedPixie*. Available at: https://www.redpixie.com/blog/iaas-paas-saas. (accessed 27/04/15)

Jakobsson, M., Shi, E., Golle, P. and Chow, R., 2009, August. Implicit authentication for mobile devices. In *Proceedings of the 4th USENIX conference on Hot topics in security* (pp. 9-9).

Kanna, T.R., Nagaraju, M. and Bhaskar, C.V., 2015. Secure Fog Computing: Providing Data Security. *IJRCCT*, *4*(1), pp.053-055.

Karame, G.O. and Stavrou, A., 2017, October. CCSW'17: 2017 ACM Cloud Computing Security. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (pp. 2627-2628). ACM.

Kashyap, A., Kumar, G.S., Jangir, S., Pilli, E.S. and Mishra, P., 2017, September. IHIDS: Introspection-based hybrid intrusion detection system in cloud environment. In *Advances in Computing, Communications and Informatics (ICACCI), 2017 International Conference on* (pp. 687-693). IEEE.

Bakar, K.A.A. and Haron, G.R., 2014, August. Adaptive authentication based on analysis of user behavior. In *Science and Information Conference (SAI), 2014* (pp. 601-606). IEEE.

Khalil, I.M., Khreishah, A. and Azeem, M., 2014. Cloud computing security: A survey. *Computers*, *3*(1), pp.1-35.

Krishnan, D. and Chatterjee, M., 2012, October. An adaptive distributed intrusion detection system for cloud computing framework. In *International Conference on Security in Computer Networks and Distributed Systems* (pp. 466-473). Springer, Berlin, Heidelberg.

Kumar, S., Sim, T., Janakiraman, R. and Zhang, S., 2005, December. Using continuous biometric verification to protect interactive login sessions. In *Computer Security Applications Conference, 21st Annual* (pp. 10-pp). IEEE.

Le, C. and Jain, R., 2011. A survey of biometrics security systems. *EEUU. Washington University in St. Louis*.

Li-qin,T., Lin, C. and Ni, Y., 2010, October. Evaluation of user behavior trust in cloud computing. In *Computer Application and System Modeling (ICCASM), 2010 International Conference on* (Vol. 7, pp. V7-567). IEEE.

Li F., 2012. Behaviour Profiling for Mobile Devices. Plymouth.

Li, F., Clarke, N., Papadaki, M. and Dowland, P., 2011. Misuse detection for mobile devices using behaviour profiling. *International Journal of Cyber Warfare and Terrorism (IJCWT)*, *1*(1), pp.41-53.

Li, F., Clarke, N., Papadaki, M. and Dowland, P., 2014. Active authentication for mobile devices utilising behaviour profiling. *International journal of information security*, *13*(3), pp.229-244.

Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J. and Ghalsasi, A., 2011. Cloud computing—The business perspective. *Decision support systems*, *51*(1), pp.176-189.

Mather, T., Kumaraswamy, S. and Latif, S., 2009. *Cloud security and privacy: an enterprise perspective on risks and compliance*. " O'Reilly Media, Inc.".

Meena, S. and Syal, R., 2017, February. Authentication scheme in cloud computing: A review. In *Electrical, Computer and Communication Technologies (ICECCT), 2017 Second International Conference on* (pp. 1-6). IEEE.

Mell Peter and grance T., 2011. The NIST Definition of Cloud Computing Recommendations of the National Institute of Standards and Technology. *Nist Special Publication* 145: 7.

Meng FJ, Zhang X, Chen P and Xu JM.,2017. DriftInsight: Detecting Anomalous Behaviors in Large-Scale Cloud Platform. *IEEE International Conference on Cloud Computing, CLOUD* 2017-June: 230-237.

Moreau, Y., Verrelst, H. and Vandewalle, J., 1997, October. Detection of mobile phone fraud using supervised neural networks: A first prototype. In *International Conference on Artificial Neural Networks* (pp. 1065-1070). Springer, Berlin, Heidelberg.

Mushtaq, N., Werner, P., Tolle, K. and Zicari, R., 2004, July. Building and evaluating nonobvious user profiles for visitors of Web sites. In *e-Commerce Technology, 2004. CEC 2004. Proceedings. IEEE International Conference on* (pp. 9-15). IEEE.

Nanavati, S., Thieme, M. and Nanavati, R., 2002. *Biometrics: identity verification in a networked world* (Vol. 20). John Wiley & Sons.

Ogwueleka, F.N., 2009. Fraud detection in mobile communications networks using user profiling and classification techniques. *Journal of Science and Technology (Ghana)*, *29*(3).

Pandeeswari, N. and Kumar, G., 2016. Anomaly detection system in cloud environment using fuzzy clustering based ANN. *Mobile Networks and Applications*, *21*(3), pp.494-505.

Peoples, C., Parr, G., Scotney, B., Sarangi, S. and Kar, S., 2014, September. Profiling user behaviour for efficient and resilient cloud management. In *Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on* (pp. 2636-2642). IEEE.

Prabhakar, S., Pankanti, S. and Jain, A.K., 2003. Biometric recognition: Security and privacy concerns. *IEEE security & privacy*, (2), pp.33-42.

Prasanth, A., Bajpei, M., Shrivastava, V. and Mishra, R.G., 2015. Cloud computing: A survey of associated services. *Book chapter of cloud computing: Reviews, surveys, tools, techniques and applications-an open-access eBook published by HCTL open*.

Qayyum, S., Mansoor, S., Khalid, A., Halim, Z. and Baig, A.R., 2010, June. Fraudulent call detection for mobile networks. In *Information and Emerging Technologies (ICIET), 2010 International Conference on* (pp. 1-5). IEEE.

Ross, A. and Jain, A.K., 2004, September. Multimodal biometrics: An overview. In *Signal Processing Conference, 2004 12th European* (pp. 1221-1224). IEEE.

Eludiora, S., Abiona, O., Oluwatope, A., Oluwaranti, A., Onime, C. and Kehinde, L., 2011. A user identity management protocol for cloud computing paradigm. *International Journal of Communications, Network and System Sciences*, *4*(03), p.152.

Salem, M.B. and Stolfo, S.J., 2011, September. Modeling user search behavior for masquerade detection. In *International Workshop on Recent Advances in Intrusion Detection* (pp. 181-200). Springer, Berlin, Heidelberg.

Samfat, D. and Molva, R., 1997. IDAMN: an intrusion detection architecture for mobile networks. *IEEE Journal on Selected Areas in Communications*, *15*(7), pp.1373-1380.

Sarker, A., Ginn, R., Nikfarjam, A., O'Connor, K., Smith, K., Jayaraman, S., Upadhaya, T. and Gonzalez, G., 2015. Utilizing social media data for pharmacovigilance: a review. *Journal of biomedical informatics*, *54*, pp.202-212.

Shaukat, U., Ahmed, E., Anwar, Z. and Xia, F., 2016. Cloudlet deployment in local wireless networks: Motivation, architectures, applications, and open challenges. *Journal of Network and Computer Applications*, *62*, pp.18-40.

Shi, E., Niu, Y., Jakobsson, M. and Chow, R., 2011, October. Implicit authentication through learning user behavior. In *International Conference on Information Security* (pp. 99-113). Springer, Berlin, Heidelberg.

SmartDataCollective, 2013. *Companies Who Have Moved to the Cloud | SmartData Collective*. . Available at: http://smartdatacollective.com/gilallouche/145341/7-well-known-companies-have-moved-cloud (accessed 27/04/15).

Sola, J. and Sevilla, J., 1997. Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Transactions on Nuclear Science*, *44*(3), pp.1464-1468.

Spanaki, P. and Sklavos, N., 2018. Cloud Computing: Security Issues and Establishing Virtual Cloud Environment via Vagrant to Secure Cloud Hosts. In *Computer and Network Security Essentials* (pp. 539-553). Springer, Cham.

Stolfo, S.J., Salem, M.B. and Keromytis, A.D., 2012, May. Fog computing:

Mitigating insider data theft attacks in the cloud. In *Security and Privacy Workshops (SPW), 2012 IEEE Symposium on* (pp. 125-128). IEEE.

Subudhi, S. and Panigrahi, S., 2015. Quarter-sphere support vector machine for fraud detection in mobile telecommunication networks. *Procedia Computer Science*, *48*, pp.353-359.

Sudha, I., Kannaki, A. and Jeevidha, S., 2014. Alleviating internal data theft attacks by decoy technology in cloud. *IJCSMC, March*.

Sun, B., Yu, F., Wu, K. and Leung, V., 2004, October. Mobility-based anomaly detection in cellular mobile networks. In *Proceedings of the 3rd ACM workshop on Wireless security*(pp. 61-69). ACM.

Sun, B., Chen, Z., Wang, R., Yu, F. and Leung, V.C., 2006, January. Towards adaptive anomaly detection in cellular mobile networks. In *Consumer Communications and Networking Conference, 2006. CCNC 2006. 3rd IEEE* (Vol. 2, pp. 666-670). IEEE.

Tiwari, T., Turk, A., Oprea, A., Olcoz, K. and Coskun, A.K., 2017, December. User-profile-based analytics for detecting cloud security breaches. In *Big Data (Big Data), 2017 IEEE International Conference on* (pp. 4529-4535). IEEE.

Vaquero, L.M., Rodero-Merino, L., Caceres, J. and Lindner, M., 2008. A break in the clouds: towards a cloud definition. *ACM SIGCOMM Computer Communication Review*, *39*(1), pp.50-55.

Violina B.,2018. *12 top cloud security threats for 2018: The dirty dozen | CSO Online*. *CSO Online*. Available at: https://www.csoonline.com/article/3043030/security/12-top-cloud-security-threats-for-2018.html (accessed 10/08/18).

Virvilis, N. and Gritzalis, D., 2013, September. The big four-what we did wrong in advanced persistent threat detection?. In *Availability, Reliability and Security (ARES), 2013 Eighth International Conference on* (pp. 248-254). IEEE.

Wang, L. ed., 2010. *Behavioral Biometrics for Human Identification: Intelligent Applications: Intelligent Applications*. IGI Global.

Wood T, Cecchet E, Ramakrishnan KK, Shenoy P, Merwe J Van Der and Venkataramani A.,2010. Disaster recovery as a cloud service : Economic benefits & beployment challenges University of Massachusetts Amherst. *Proceeding of the 2nd USENIX Workshop on Hot Topics in Cloud Computing* 1–7.

Woodward Jr, J.D., Horn, C., Gatune, J. and Thomas, A., 2003. *Biometrics: A look at facial recognition*. RAND CORP SANTA MONICA CA.

Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N.N., Kunin, V., Goodwin, L., Wu, M., Tindall, B.J. and Hooper, S.D., 2009. CLOUD COMPUTING–An Overview An Overview,‖. *White Pap*, *462*(7276), pp.1-5.

Yampolskiy, R.V., 2008. Behavioral modeling: an overview. *American Journal of Applied Sciences*, *5*(5), pp.496-503.

Yampolskiy, R.V. and Govindaraju, V., 2008. Behavioural biometrics: a survey and classification. *International Journal of Biometrics*, *1*(1), pp.81-113.

Yan, Q., Yu, F.R., Gong, Q. and Li, J., 2016. Software-defined networking (SDN) and distributed denial of service (DDoS) attacks in cloud computing environments: A survey, some research issues, and challenges. *IEEE Communications Surveys & Tutorials*, *18*(1), pp.602-622.

Yang, Y.C., 2010. Web user behavioral profiling for user identification. *Decision Support Systems*, *49*(3), pp.261-271.

Yazji, S., Chen, X., Dick, R.P. and Scheuermann, P., 2009, July. Implicit user re-authentication for mobile devices. In *International Conference on Ubiquitous Intelligence and Computing* (pp. 325-339). Springer, Berlin, Heidelberg.

Yazji, S., Dick, R.P., Scheuermann, P. and Trajcevski, G., 2011, September. Protecting Private Data on Mobile Systems based on Spatio-temporal Analysis. In *PECCS* (pp. 114-123).

Yazji, S., Scheuermann, P., Dick, R.P., Trajcevski, G. and Jin, R., 2014. Efficient location aware intrusion detection to protect mobile devices. *Personal and Ubiquitous Computing*, *18*(1), pp.143-162.

Ye, X., Chen, X., Wang, H., Zeng, X., Shao, G., Yin, X. and Xu, C., 2016. An anomalous behavior detection model in cloud computing. *Tsinghua Science and Technology*, *21*(3), pp.322-332.

Yeung, D.Y. and Ding, Y., 2003. Host-based intrusion detection using dynamic and static behavioral models. *Pattern recognition*, *36*(1), pp.229-243.

Zhang, J. and Shukla, M., 2006, December. Rule-based platform for web user profiling. In *Data Mining, 2006. ICDM'06. Sixth International Conference on* (pp. 1183-1187). IEEE.

Zhang, Q., Cheng, L. and Boutaba, R., 2010. Cloud computing: state-of-the-art and research challenges. *Journal of internet services and applications*, *1*(1), pp.7-18.

Zissis, D. and Lekkas, D., 2012. Addressing cloud computing security issues. *Future Generation computer systems*, *28*(3), pp.583-592.

Publication.

1. Al-Bayati, B., Clarke, N. and Dowland, P., 2016. Adaptive Behavioral Pro-filing for Identity Verification in Cloud Computing: A Model and Prelimi-nary Analysis. GSTF Journal on Computing (JoC), ISSN: 2251-3043, Vol. 5, Iss.1, pp21-28, 2016.

2. Al-Bayati, B., Clark, N., Haskell-Dowland, P. and Li, F., 2018, June. Con-tinuous identity verification in cloud storage services using behavioural profiling. In 17th European Conference on Cyber Warfare and Security, ISBN: 978-1-911218-85-2, pp1-10. Academic Conferences and Publish-ing International Limited.

3. Al-Bayati, B., Clark, N., Haskell-Dowland, P. and Li, F., 2018. Misuse De-tection in a Simulated IaaS Environment (in progress). It is accepted by Emerging Technologies for Authorization and Authentication (ETAA) 2018, and it will be published by Springer Verlag.

Appendix A – Ethical Approval, Consent Form and Information Sheet (Data Collection SaaS)

RESEARCH WITH PLYMOUTH UNIVERSITY

15 July 2016

CONFIDENTIAL
Burhan Al-Bayati
School of Computing, Electronics and Mathematics

Dear Burhan

Ethical Approval Application

Thank you for submitting the ethical approval form and details concerning your project:

Behavioural profiling for users' Dropbox usage

I am pleased to inform you that this has been approved.

Kind regards

Paula Simson
Secretary to Faculty Research Ethics Committee

Cc. Prof Nathan Clarke
Prof Paul Dowland
Dr Fudong Li

Faculty of Science and Engineering
Plymouth University
Drake Circus
PL4 8AA

T +44 (0) 1752 584 584
F +44 (0) 1752 584 540
W www.plymouth.ac.uk

Mrs Christine Mushens BA
Faculty Business Manager

SAMPLE SELF-CONSENT FORM

PLYMOUTH UNIVERSITY

FACULTY OF SCIENCE AND ENGINEERING

Human Ethics Committee Sample Consent Form

CONSENT TO PARICIPATE IN RESEARCH PROJECT / PRACTICAL STUDY

Name of Principal Investigator

Burhan Al-bayati

Title of Research

Behavioural profiling for users' Dropbox usage

Brief statement of purpose of work

There is no doubt that the flexible and convenient facilities of cloud computing services have changed our daily lives (whether people are aware of it or not); however, the biggest barrier that hinders the development and widespread use of cloud computing services are the security issues one of which is being merely at the point-of-entry. In order to secure sensitive services, it is imperative to increase the level of authentication beyond the standard point-of-entry technique. Although a number of research has been undertaken exploring this, they were merely a theoretical proposal without any practical experiment, or evaluated based upon simulated data.

This study seeks to investigate and specify the appropriate user's behaviour activities that need to be deployed and incorporated in the proposed solution to provide a more secure and user-friendly.

Therefore, this experiment seeks to collect a real data of the historical usage of Dropbox users, in order to evaluate the appropriateness and effectiveness of utilising them for such continuous identity verification.

As a participant, no modification will be made upon your computer as your Dropbox activities will be collected from the web browser. Please merely use your Dropbox as you normally do. Your historical usage for last 3 months will be collected without any interference upon your normal activities. The files will be taken by the principal investigator and will be anonymous and stored in a secure location within the Centre for Security, Communications and Network Research (CSCAN) at Plymouth University.

SAMPLE INFORMATION SHEET FOR ADULT / CHILD

PLYMOUTH UNIVERSITY
FACULTY OF SCIENCE AND ENGINEERING

RESEARCH INFORMATION SHEET

Name of Principal Investigator

Burhan Al-Bayati

Title of Research

Behavioural profiling for users' Dropbox usage

Aim of research

This study seeks to investigate and specify the appropriate user's behaviour activities of Dropbox that need to be deployed and incorporated in the proposed solution to provide a more secure and user-friendly. Therefore, this experiment seeks to collect a real data of the historical usage of Dropbox users, in order to evaluate the appropriateness and effectiveness of utilising them for such continuous identity verification.

Description of procedure

The participants will copy their historical activities for the last 3 months and store them in Excel file, then send them to the researcher after removing any information that related to privacy such as file names

Description of risks

At no stage will any personally identifiable information be seen by any individual neither the researchers nor on any publication. The captured data will be stored after being converted to measurement features. All of the information will be treated confidentially and data will be anonymous, storage and publication of research material.

Benefits of proposed research

The ultimate aim of this research project is to build continuous identity verification for cloud computing services. A verification system built upon this would provide a more secure and user-friendly.

Right to withdraw

You have the right to withdraw at any stage without giving a reason. Your dataset will be removed and securely deleted.

Contact for Further Information

If you are dissatisfied with the way the research is conducted, please contact the principal investigator in the first instance: Burhan Al-Bayati, A317, Portland Square Building, Plymouth University. Email: burhan.al-bayati@plymouth.ac.uk, Telephone number *[+441752586259]*. If you feel the problem has not been resolved please contact the secretary to the Faculty of Science and Engineering Human Ethics Committee: Mrs Paula Simson 01752 584503.

At all stages of the study, confidentiality of the collected data and subsequent analysis will be maintained. At no time, will any identifying information about the participants be used in any publication or research output.

For information regarding the study, please contact:

Burhan Al-bayati - burhan.al-bayati @plymouth.ac.uk
For any questions concerning the ethical status of this study, please contact the secretary of the Human Ethics Committee – paula.simson@plymouth.ac.uk

---

The objectives of this research have been explained to me.

I understand that I am free to withdraw from the research at any stage, and ask for my data to be destroyed if I wish.

I understand that my anonymity is guaranteed, unless I expressly state otherwise.

I understand that the Principal Investigator of this work will have attempted, as far as possible, to avoid any risks, and that safety and health risks will have been separately assessed by appropriate authorities (e.g. under COSHH regulations).

Under these circumstances, I confirm that I am 18 years old or above and I agree to participate in the research.

Name:.....

Signature:..... Date:.....

Appendix B – Ethical Approval, Consent Form and Information Sheet (Data Collection IaaS)

RESEARCH WITH PLYMOUTH UNIVERSITY

23 October 2017

CONFIDENTIAL
Burhan Al-Bayati
School of Computing, Electronics and Mathematics

Dear Burhan

Ethical Approval Application

Thank you for submitting the ethical approval form and details concerning your project:

Behavioural Profiling of Users

I am pleased to inform you this has been approved.

Kind regards

Paula Simson
Secretary to Faculty Research Ethics Committee

Cc. Prof Nathan Clarke

PLYMOUTH UNIVERSITY

FACULTY OF SCIENCE AND ENGINEERING

Human Ethics Committee Consent Form

CONSENT TO PARICIPATE IN RESEARCH PROJECT / PRACTICAL STUDY

Name of Principal Investigator
Burhan Al-Bayati

Title of Research
Behavioural Profiling of Users

Brief statement of purpose of work

The aim of the project is to develop a real-time service utilisation monitor, that is a piece of software which in real time provides analysis (based on the classifications from collected data) of the usage of the system and tries to determine if the user should be authenticated or not. i.e. is the correct user operating the system. The premise of the technique is based on user's behavioural patterns with software applications. For example, users tend to have a defined pattern of things to do when they first enter the office – open Internet Explorer to read the news, catch up on emails, check their schedule for day etc.

The aim of this study is to collect data on this usage from a number of sample users and from this identify discriminative features which can differentiate one user from another.

The objectives of this research have been explained to me.

I understand that I am free to withdraw from the research at any stage, and ask for my data to be destroyed if I wish.

I understand that my anonymity is guaranteed, unless I expressly state otherwise.

I understand that the Principal Investigator of this work will have attempted, as far as possible, to avoid any risks, and that safety and health risks will have been separately assessed by appropriate authorities.

Under these circumstances, I agree to participate in the research.

Name: ……

Signature: ……. Date: ……..

PLYMOUTH UNIVERSITY

FACULTY OF SCIENCE AND ENGINEERING

Human Ethics Committee Information Sheet.

Name of Principal Investigator

Burhan Al-Bayati

Title of Research

Behavioural Profiling of Users

Aim of research

This study seeks to investigate the use of a user's behaviour activities during interacting with the computer to provide a more secure and user-friendly authentication approach. As such, this experiment seeks to collect real data from the users' interactions with their computers, in order to evaluate the appropriateness and effectiveness of utilising them for such a continuous identity verification.

Description of procedure

The software will be installed by investigator on the participants' computers for two-week period, will be limited to:
1)Application usage data – time and date stamps of current program/application open, close and 'focus' events.

2)Domains of websites you visit with time stamps – although this is cryptographically hashed to maintain privacy (i.e. the domains are not stored in clear text and cannot be easy to identify the website).

3) Your name will not be connected to your data, your name will only be taken with regards to your consent to take part in the trial.

Description of risks

At no stage will any personally identifiable information be seen by any individual neither the researchers nor on any publication. The captured data will be stored after being converted to measurement features. All of the information will be treated confidentially and data will be anonymous, storage and publication of research material.

Benefits of proposed research

The ultimate aim of this research project is to build continuous identity verification system. A verification system built upon this would provide a more secure and user-friendly.

Right to withdraw

Your data is not collected from your machine until the end of the trial period, as such the data until the collection time is stored solely in a local file on your computer. If you decide to withdraw from this trial, you have the right to withdraw at any stage without giving a reason. Your dataset will be removed and securely deleted.

Contact for Further Information

If you are dissatisfied with the way the research is conducted, please contact the principal investigator in the first instance: Burhan Al-Bayati, A317, Portland Square Building, Plymouth University. Email: burhan.al-bayati@plymouth.ac.uk, Telephone number [*07459012291]*. If you feel the problem has not been resolved please contact the secretary to the Faculty of Science and Engineering Human Ethics Committee: Mrs Paula Simson 01752 584503.